THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XVI | March, 1925 | Number 3

AN EVALUATION OF CERTAIN INCENTIVES USED IN SCHOOL WORK

ELIZABETH B. HURLOCK

University Extension, Columbia University

Contrary to the belief that praise is a more effective form of incentive than reproof to use in the case of school children, the writer, in an experimental study[1] found that after one application the two, on the whole, were of equal value. These rather surprising results at once opened up the question of whether reproof, if used over a given period of time, would not lose some of its effectiveness. Previous studies, using the practice experiment technique with children as subjects, have shown that encouragement brought about increasingly better work throughout the period of practice, but no such studies are available to give us information in regard to the effect of the continuous use of discouragement. Kirby,[2] used 1350 New York school children from the Grades III and IV classes to determine the value of encouragement upon their performance in addition. He had each child practice adding for 75 minutes, but this time was distributed differently for the different groups. The form of encouragement used was the competition of each child with the record he made on the previous day's performance, and not rivalry with the other children of the group. The results obtained showed a median gain of 48 per cent, but as no control group was used, we have no means of determining how much of this was due to encouragement, and how much to practice.

---

[1] Hurlock, E. B.: The Value of Praise and Reproof as Incentives for Children. *Archives of Psychology*, No. 71, New York, July, 1924.

[2] Kirby, Thomas J.: Practice in the Case of School Children. *College University Contributions to Education*, Teachers College Series, No. 58.

In 1917, Chapman and Feder[1] made a similar study, but introduced a control group to check up on the effect of practice. Three tests, the Thorndike Simple Addition Test, the Woodworth-Wells Cancellation Test, and the Digit-symbol Test, were given to 36 boys and girls of Grade V*a* of the Cleveland Observation School. One half of this group was motivated by incentives similar to those used by Kirby, and the other half was tested separately and received no incentives at all. The tests were given over a period of 10 days; 10 minutes each day was devoted to addition, 5 to substitution, and 1 to cancellation. The results for all three tests showed a decided improvement in the case of the motivated group. The average scores for the 1st, 5th and 10th days were 14.7, 18.6 and 20.2 for the motivated, and 14.0, 11.7, and 11.9 for the non-motivated group in addition: 81, 155, and 215 for the motivated and 79, 130, and 181 for the non-motivated in substitution: and 32.0, 46.8, and 55.2 for the motivated and 34.0, 47.9, and 51.2 for the non-motivated in cancellation.

In the present experiment, an attempt has been made to evaluate reproof, when used continuously over a given period of time, with praise used in a similar manner. In addition to this, the writer endeavored to discover what, if any, improvement would be found in the work of a group of children who, although present when the others received encouragement and discouragement, was completely ignored at every practice period. The problem, then, stated briefly is this: In a classroom, do the children who constantly receive praise for their work show more improvement from day to day than do the children who are reproved or who are completely ignored?

The 106 children used in this study were from the classes of Grades IV and VI of the Cameron Public School Building of Harrisburg, Penna.[2] Forty-eight boys and 60 girls formed these two groups. The average ages of the children are shown in Table I. From this, it may be seen that the average age of Grade VI children is 11.92, and of Grade IV children, 9.37 years. In Grade VI, the girls are, on the

[1] Chapman, J. Crosby and Feder, R. B.: The Effect of External Incentives on Improvement. *Journal of Educational Psychology*, 1917, Vol. VIII, pp. 469–474.

[2] The writer is deeply indebted to Dr. Garwood, Superintendent of the Harrisburg Public Schools, and to Mr. Getz, Principal of the Cameron School, for their kind cooperation and assistance in the carrying out of this experiment. She is also very grateful to Miss Mary Kunkel of Harrisburg for her aid in the conduct of the experiment.

average, slightly older than the boys, and in Grade IV, the opposite is true.

TABLE I.—CHRONOLOGICAL AGES OF THE SUBJECTS

	Averages			σ			PE		
	Boys	Girls	Combined	Boys	Girls	Combined	Boys	Girls	Combined
Grade VI.....	11.72	12.11	11.92	1.07	1.29	1.19	.15	.18	.12
Grade IV.....	9.41	9.33	9.37	.91	1.04	1.01	.13	.12	.09

The tests used to measure the rate of improvement under the different conditions of practice were modifications, made by the writer, of the Courtis Research Tests in arithmetic, addition form. Each test contained 30 examples of equal difficulty, each of which was made up of 6 three-place numbers. Fifteen minute periods were allowed each day for the practice. A different test was used on each of the five days.

An attempt was made to have the procedure as nearly like that of the ordinary school procedure as possible, but at the same time to control it so that the results would be scientifically reliable. All of the children, except the Control Group, were given the tests in the same room, instead of in separate rooms, as was done by previous investigators. The Praised Group received praise, and the Reproved Group, reproof, before the other members of the class. The Ignored Group heard these comments, but they themselves received no incentive whatsoever. The Control Group, after the first test, was tested in a separate room.

A detailed description of the procedure would be impossible within the limits of this paper. However, the main points may be indicated briefly. On the first day of the experiment, all of the children within each grade were given the addition test together. On the basis of these results, four equivalent groups were formed, to be known throughout the experiment as the Control, Praised, Reproved, and Ignored Group. These groups were equal not only in initial ability as displayed on these tests in addition, but also in average age and number of boys and girls within each group. The Control Group, after the first test, was separated from the others, and was given the tests on the four remaining days in a separate room. No comment was made to them: They were merely asked to add the examples on

the paper given to them at the beginning of the period. The other three groups took these tests together, but under different conditions of incentive. Each day, before the test papers were given out, the names of the children of the Praised Group were called out by the experimenter, and the children were asked to rise, come to the front of the room and face the class. They were then praised for the excellency of their work of the preceding day, as shown both in the improvement made and the general superiority over the other members of the group. They were encouraged to do better, to try to avoid any careless mistakes, and to add as many problems as the time permitted. Following this, the members of the Reproved Group were called out and were severely reproved for poor work, careless mistakes, lack of improvement over the preceding day's work, and general inferiority to the other members of the class. The members of the Ignored Group heard the praise and reproof given to the two other groups, but they themselves received no recognition whatsoever.

TABLE II.—DISTRIBUTION OF CHILDREN

Groups	Grade VI			Grade IV			Totals
	Boys	Girls	Combined	Boys	Girls	Combined	
Control	6	6	12	5	9	14	26
Praised	6	6	12	6	9	15	27
Reproved	6	6	12	6	9	15	27
Ignored	6	6	12	5	9	14	26
Totals	24	24	48	22	36	58	106

Table II shows the distribution of children into the different groups within Grades IV and VI. An equal number of boys and of girls, it may be seen, was placed in each group.

The total results are presented in Table III. A consideration of these shows some interesting facts. The four groups are approximately equal in average and variability scores on the first test, but differences begin to appear in the second test and are shown throughout the series. The greatest improvement shown at the end of the series of tests occurred in the Praised Group, where the average increase

TABLE III.—AVERAGE SCORES. TOTAL RESULTS

Groups	Averages					σ					PE$_{AV.}$				
	First day	Second day	Third day	Fourth day	Fifth day	First day	Second day	Third day	Fourth day	Fifth day	First day	Second day	Third day	Fourth day	Fifth day
Control	11.81	12.34	11.65	10.50	11.35	5.53	5.25	5.81	5.11	4.21	.73	.69	.76	.67	.55
Praised	11.81	16.59	18.85	18.81	20.22	6.17	7.16	7.11	7.53	7.68	.79	.92	.91	.97	.99
Reproved	11.85	16.59	14.30	13.26	14.19	6.55	6.84	6.97	6.54	6.78	.84	.88	.89	.84	.87
Ignored	11.84	14.19	13.30	12.92	12.38	5.64	6.30	6.05	6.33	6.06	.74	.83	.79	.83	.79

RELIABILITY OF DIFFERENCES—COMPARISON OF FINAL TESTS

GROUPS	ACTUAL DIFFERENCE	PE$_{DIFF.}$	DIFFERENCE IN PE UNITS	CHANCES THAT TRUE DIFFERENCE IS ABOVE 0
Praised—control	8.87	1.13	7.85	10,000 in 10,000
Reproved—control	2.84	1.02	2.78	9,382 in 10,000
Ignored—control	1.03	.95	1.08	5,338 in 10,000

was 8.41; the next largest in the Reproved Group, 2.34; and then only .54 in the Ignored Group, and a loss of .46 points on the average in the Control Group. When the statistical reliability of the differences was computed, it was found that only in the case of the Praised Group was the gain in the fifth test, over and above gain due merely to practice, as measured by the Control Group, large enough to be

TABLE IV.—PERCENTAGE OF AVERAGE INCREASE WITHIN EACH GROUP. TOTAL RESULTS

Groups	Third over first test, per cent	Third test over control, per cent	Fifth over first test, per cent	Fifth test over control, per cent
Control	−1	..	−4	
Praised	59	60	71	75
Reproved	21	22	19	23
Ignored	12	13	5	9

reliable. In the Reproved and Ignored Groups, gains did occur, but these were only indications of true differences, and were not large enough for absolute certainty. In like manner, a study of the percentage of average improvement within each group (results presented in Table II) shows that the Praised Group is the only one which made a decided gain. The Reproved and Ignored Groups showed slightly greater gain in the third over the first than in the fifth over the first, while the Control Group showed a slight amount of decrease both times. The Praised Group is the only one which shows an increase in proficiency from the beginning of the series to the end.

A comparison of the boys and girls within the four groups was then made. The results are summarized in Tables V and VI. Table V shows the average gains for the two sexes taken separately, together with the σ and PE scores. It may be seen that in the Control Group, neither sex shows much gain throughout the series. In the last test, both do, on the average, slightly worse than on the first test. The results of the Praised Group show for both sexes a decided increase at first, followed throughout by a more gradual one as the physiological limit is being reached. The girls, however, showed a greater increase than did the boys. After the first administration of reproof, the boys

TABLE V.—AVERAGE SCORES. SEX COMPARISONS

Groups	Averages					σ					$PE_{AV.}$				
	First day	Second day	Third day	Fourth day	Fifth day	First day	Second day	Third day	Fourth day	Fifth day	First day	Second day	Third day	Fourth day	Fifth day
Control															
Boys	13.45	13.91	13.18	12.18	11.81	5.73	5.10	5.43	4.69	3.72	1.16	1.03	1.09	.94	.75
Girls	10.60	11.20	10.54	9.26	10.33	5.09	5.11	5.66	5.02	4.31	.87	.88	.98	.86	.74
Praised															
Boys	13.16	17.34	19.83	19.00	20.58	6.84	6.80	7.23	7.52	7.71	1.32	1.31	1.40	1.46	1.49
Girls	10.73	16.00	18.06	18.66	19.93	5.70	7.18	7.37	7.54	7.66	.98	1.24	1.27	1.56	1.32
Reproved															
Boys	13.25	18.08	17.00	15.16	16.66	7.44	6.60	6.94	6.39	7.45	1.44	1.28	1.34	1.23	1.44
Girls	10.73	15.40	12.13	11.73	12.20	5.50	6.59	5.45	6.02	5.60	.95	1.13	.94	1.04	.96
Ignored															
Boys	13.54	16.73	17.25	14.00	13.82	5.82	6.00	6.18	6.47	6.42	1.18	1.21	1.25	1.31	1.29
Girls	10.60	12.33	12.46	12.13	11.33	5.17	6.56	6.57	6.18	5.52	.89	1.13	1.13	1.06	.95

did slightly better than following praise, and the girls slightly worse. With continued use, reproof seemed to lose some of its effectiveness. Contrary to the results found for the Praised Group, it served as a more powerful incentive for the boys throughout the series than for the girls, but at no time after the first application did it bring about work comparable to praise. Perhaps the most interesting sex difference was found in the case of the Ignored Group. On the second trial, the boys did almost as well as when some incentive was used. At the end of the series, there was almost no change in the results as compared with the first test. The girls did not show such variations. They did slightly better at first than did the members of the Control Group, but later a decrease in output suggested a decrease in effort. Table VI gives the percentage of average increase within the groups in the third and fifth trials above the first, and above practice as measured by the Control Group.

TABLE VI.—PERCENTAGE OF AVERAGE INCREASE WITHIN EACH GROUP. SEX COMPARISONS

Groups	Third over first test, per cent	Third test over control, per cent	Fifth over first test, per cent	Fifth test over control, per cent
Control				
Boys	−2	..	−12	
Girls	−1	..	− 2	
Praised				
Boys	43	45	56	68
Girls	68	69	86	88
Reproved				
Boys	28	30	26	38
Girls	13	14	14	16
Ignored				
Boys	27	29	2	14
Girls	17	18	6	8

Inasmuch as there were a great many age levels represented within the two grades, it was thought advisable to treat the age comparisons as grade comparisons, and to allow the average age of each grade to represent the age level of that group. Referring to Table I in which these ages are given, we find that 12 years is the average age for the Grade VI children, and 9.40 for those of Grade IV. Tables VII

TABLE VII.—AVERAGE SCORES. GRADE COMPARISONS

Groups	Averages					σ					PE$_{AV.}$				
	First day	Second day	Third day	Fourth day	Fifth day	First day	Second day	Third day	Fourth day	Fifth day	First day	Second day	Third day	Fourth day	Fifth day
Control															
Grade VI	15.08	16.08	15.66	13.58	13.91	4.93	4.52	5.32	4.92	2.87	.95	.87	1.03	.95	.55
Grade IV	9.00	9.14	8.21	7.86	9.14	4.29	3.67	3.73	5.91	3.83	.77	.66	.67	1.06	.69
Praised															
Grade VI	15.25	19.33	22.42	23.50	25.16	6.73	7.08	8.09	7.40	7.39	1.30	1.49	1.56	1.43	1.43
Grade IV	9.06	14.40	16.00	15.06	16.26	4.73	5.63	5.09	5.07	5.25	.82	.97	.88	.87	.91
Reproved															
Grade VI	15.33	20.33	18.33	17.58	17.83	6.27	5.70	6.47	6.08	6.28	1.21	1.10	1.26	1.18	1.21
Grade IV	9.06	13.60	11.06	9.80	11.26	5.63	5.99	4.68	4.22	5.82	.97	1.03	.81	.75	1.00
Ignored															
Grade VI	15.16	17.99	18.16	17.58	16.58	5.18	5.56	4.12	5.27	5.69	1.00	1.08	.79	1.02	1.10
Grade IV	9.00	10.93	9.14	8.93	8.79	4.05	5.05	4.21	4.15	3.47	.72	.91	.76	.74	.62

and VIII contain the results when treated separately for the two grades. A study of Table VII reveals the fact that in the Control Group, the older children did slightly worse at the end of the series than at the beginning, and the younger ones about the same. Praise and reproof brought about increased work for both younger and older children, but of the two, praise was the decidedly more effective. In the Ignored Group, the older children did slightly better throughout the series, while the work of the younger ones remained about the

TABLE VIII.—PERCENTAGE OF AVERAGE INCREASE WITHIN EACH GROUP. GRADE COMPARISONS

Groups	Third over first test, per cent	Third test over control, per cent	Fifth over first test, per cent	Fifth test over control, per cent
Control				
Grade VI	3	..	−8	
Grade IV	−9	..	2	
Praised				
Grade VI	47	44	65	73
Grade IV	77	86	79	77
Reproved				
Grade VI	19	16	16	24
Grade IV	22	31	24	22
Ignored				
Grade VI	20	17	9	17
Grade IV	2	11	−2	00

same. Table VIII shows the percentage of average increase of these results within the group and above the practice effect. It emphasizes the slight tendency to improvement in the Control Group when no incentive was used, the remarkable increase for both groups, especially the younger, following praise, and the decidedly less increase as the result of reproof. The pupils of Grade VI, when ignored, increased their work at first, and later showed little improvement, while those of Grade IV did practically the same amount of work throughout the series.

The next question to be investigated was whether the incentives were of different degrees of effectiveness according to the initial

TABLE IX.—AVERAGE SCORES. COMPARISON OF DIFFERENT GRADES OF ABILITY

Groups	Averages					σ					PEAv.				
	First day	Second day	Third day	Fourth day	Fifth day	First day	Second day	Third day	Fourth day	Fifth day	First day	Second day	Third day	Fourth day	Fifth day
Control															
Superior	17.00	16.78	15.22	13.00	13.11	3.88	4.42	5.01	4.83	3.18	.86	.98	1.12	1.08	.71
Average	11.75	11.62	12.25	11.12	11.75	3.94	4.43	5.85	5.55	4.44	.93	1.04	1.38	1.31	1.04
Inferior	6.78	9.11	7.77	7.22	9.22	2.71	3.78	3.74	2.36	3.91	.61	.84	.83	.53	.87
Praised															
Superior	18.11	23.33	24.77	24.11	25.89	4.20	5.32	5.81	6.20	6.45	.93	1.15	1.29	1.38	1.50
Average	11.22	15.77	19.11	18.22	20.11	3.24	4.66	6.54	7.55	7.07	.72	1.04	1.46	1.68	1.57
Inferior	6.11	10.55	12.67	13.44	14.67	2.69	4.05	3.61	4.94	4.79	.60	.90	.81	1.10	1.08
Reproved															
Superior	17.44	22.44	19.67	17.78	19.78	4.51	5.48	6.99	7.13	7.38	1.01	1.22	1.56	1.59	1.64
Average	11.44	16.67	12.67	10.67	13.22	3.64	4.23	3.63	2.81	3.26	.81	.94	.81	.63	.73
Inferior	6.67	10.67	10.55	10.22	9.55	3.38	4.22	5.22	4.34	4.71	.75	.94	1.16	.97	1.05
Ignored															
Superior	17.11	17.89	15.67	13.22	14.22	3.96	6.03	5.15	6.83	6.92	.88	1.34	1.15	1.52	1.54
Average	11.75	15.37	15.00	14.75	13.50	3.92	6.96	6.06	6.04	5.68	.92	1.65	1.43	1.42	1.34
Inferior	6.67	9.44	9.44	9.89	9.55	2.80	3.36	5.16	5.08	4.08	.62	.74	1.15	1.13	.91

ability of the children in the test. More specifically, how do praise, reproof and being ignored affect superior children as compared with average and inferior children? All of the children within the two grade sections of each of the four experimental groups were divided into three groups—the highest one-third were called "superior;" the middle one-third, "average;" and the lowest one-third, "inferior." The results expressed in terms of average performance on each test of the series, together with the σ and $PE_{AV.}$ scores, are found in Table IX. The percentage of average increase in the third over the first, and the fifth over the first test, together with the percentage of gain made above that of the Control Group, is given in Table X. A study of these

TABLE X.—PERCENTAGE OF AVERAGE INCREASE WITHIN EACH GROUP. COMPARISONS OF DIFFERENT LEVELS OF ABILITY

Groups	Third over first test, per cent	Third test over control, per cent	Fifth over first test, per cent	Fifth test over control, per cent
Control				
Superior	−10	..	−20	
Average	4	..	00	
Inferior	15	..	36	
Praised				
Superior	37	47	43	63
Average	70	66	79	79
Inferior	107	92	140	104
Reproved				
Superior	13	23	14	34
Average	11	7	16	16
Inferior	53	38	43	7
Ignored				
Superior	−8	2	−13	7
Average	28	24	15	15
Inferior	42	27	43	7

tables reveals the fact that in the Control Group the scores for the "superior" children decreased throughout the experiment, those for the "average" remained about the same, while those for the "inferior" children showed increasing gain. Considering the results of the other three groups, over and above the gain due to practice, as measured by the Control Group as standard, we find the greatest gains in the

"inferior," the next greatest in the "average," and the least in the "superior" children, as the result of praise. Reproof likewise was most effective when used with "inferior" children, and of about the same value for "average" and "superior" children. Two facts of interest may be noted here. First, it brought about a decidedly less increase of work in each case than did praise, and secondly, it was more effective at the beginning of the series than at the end. Being reproved seemed to strike the "superior" children hardest. After the second test, we find a steady decline in average work, as was true in the Control Group. The "average" and "inferior" children did about the same as when reproved.

In Table XI is shown the percentage of accuracy for each group throughout the series. The average number of problems added, the average number correct, and in the case of the first, third and fifth tests, the percentage of problems correct are given. These results impress upon us the fact that the Praised and Reproved Groups are the only ones which show any gain in accuracy from the beginning to the end of the series. Of these two, the greatest gain, 16 per cent, is made in the Praised Group. The Control and Ignored Groups both showed a loss in accuracy at the end of the series. Of the two, the greatest loss, 17 per cent, was made by the Control Group.

Summary

The data presented above seem to point to the following conclusions:

1. The greatest amount of average improvement at the end of a series of tests, over and above practice, was found in the group that was praised for improvement in work, and was urged and encouraged to do better in the succeeding tests. Decidedly less improvement was found in the Reproved Group, less still in the Ignored, and none at all in the Control.

2. The results for the boys and girls within each group were compared, and they showed that in the case of the girls, the greatest improvement followed praise, and in the case of the boys, reproof. In both cases, however, praise was decidedly more effective than reproof. The boys who were ignored showed slightly more improvement than did the girls, while in the Control Group, both boys and girls did worse, on the average, at the end of the series of tests than at the beginning.

TABLE XI.—PERCENTAGE OF AVERAGE ACCURACY SHOWN THROUGHOUT THE SERIES

Groups	First day			Second day		Third day			Fourth day		Fifth day			Gain in accuracy	
	Solved	Correct	Per cent correct	Solved	Correct	Solved	Correct	Per cent correct	Solved	Correct	Solved	Correct	Per cent correct	Third day over first, per cent	Fifth day over first, per cent
Control	15.65	11.81	75	20.23	12.34	18.74	11.65	62	17.52	10.50	19.47	11.35	58	−13	−17
Praised	15.41	11.81	77	20.41	16.59	20.41	18.85	92	19.52	18.81	21.71	20.22	93	15	16
Reproved	15.62	11.85	76	18.04	10.59	15.89	14.30	89	15.52	13.26	16.92	14.19	84	13	8
Ignored	15.49	11.84	76	20.04	14.19	19.11	13.30	69	17.16	12.92	18.73	12.38	66	− 7	−10

3. Grade comparisons showed younger and older children to be about equal in responsiveness to the incentives of praise and reproof at the end of the series of tests. The Grade IV children did exactly the same, on the average, at the end of the series as at the beginning, when ignored, but those of Grade VI showed a slight improvement. In the Control Group, the older children decreased their work slightly, showing that they needed some incentive, while the younger ones did about the same.

4. Praise proved to be decidedly the most effective incentive to use in the case of all children, when classed as "superior," "average," or "inferior" on the initial test. Within the groups, the "inferior" children were most responsive to praise, and the "superior" to reproof. Being ignored brought about only slight improvement on the part of "superior" and "inferior" children, and little more on "average." Practice without some form of incentive showed the "superior" children to do, on the average, worse work at the end of the series than at the beginning, the "inferior" to do about one-third better, and the average exactly the same.

5. Increased accuracy in performance was found only in the work of the children who were praised or reproved. The group that was praised, however, made greater improvement than the group that was reproved. The Ignored and Control Groups both showed a decrease in accuracy at the end of the series.

These results as a whole point conclusively to the fact that regardless of the factors of age, sex, initial ability, or accuracy, praise is decidedly the most effective of the three incentives here investigated. Reproof, when first used, seemed to be about equal in value to praise, but with continued use its effectiveness showed a decided decline. To ignore children in a group where the other members are receiving some incentive seems to be psychologically bad. The longer they are ignored, the less improvement do they show, even in spite of the opportunity to improve which comes from continued practice in one form of work. Hence, we may conclude that praise is the best form of incentive to use, no matter from what angle we may regard the matter.

LATIN PROGNOSIS: A STUDY OF THE DETAILED FACTORS OF INDIVIDUAL PUPILS

ORLIE M. CLEM

Teachers College, Syracuse University

To reduce misdirected effort is the first aim of any study in educational prognosis. The method, if it is to be other than speculative, requires an analysis of the factors which have made for success in a given situation with a view to determining the probable effect of the same factors in a second situation. Assuming that specific abilities are required in different types of learning, the problem of the investigator is to devise means for segregating and measuring these specific abilities as a basis for prognosis.

The present study in prognosis has been preceded by a number of important ones. In 1914, Dr. Truman Lee Kelley determined the relative predictive value of elementary school marks, teachers' estimates, and some special tests, upon a pupil's success in mathematics, English, and history. In 1918, Dr. Agnes Low Rogers developed a group of six tests for predicting ability in mathematics. In the same year Dr. Elbert K. Fretwell ascertained that success in the first year of the junior high school could be predicted more successfuly by a group of standardized educational tests, than by either elementary school marks, or teachers' estimates, or age.

The present study was prompted by that of Dr. William Sims Allen.[1] Dr. Allen, at the beginning of the first semester, 1921–22, gave 21 psychological tests to 364 boys taking first year Latin in the Boys' High School, Brooklyn. At the end of the semester, he gave to the same pupils 11 pairs of Latin tests devised by Professor Thomas H. Briggs. These tests were constructed so that they could be objectively scored, and covered the 11 types of work done in the first semester of Latin. Dr. Allen, through multiple correlation procedure, chose from the 21 psychological tests a prognosis battery of six tests which gave the highest correlation with the Latin tests used as criteria. These six tests: Briggs' Analogies—Alpha and Beta; Thorndike Tests of Word Knowledge—A and B; Rogers' Interpolation—1 and 2, when combined, gave a multiple correlation of .588.

---

[1] Allen, William Sims: A Study in Latin Prognosis. *Teachers College Contributions to Education*, No. 135, New York, 1923.

As stated above, Allen used one factor as a basis for prognosis—psychological tests. The purpose of the present study was to find the effect of several detailed factors, including the Allen Battery of six tests, upon a pupil's success in first year Latin; to choose the most effective factors, and through multiple correlation, to obtain their combined effect as a basis for prognosis.

This study was conducted, 1922–23, under the guidance of the department of secondary education of Teachers College. It was made possible through the kindly cooperation of the classical departments of three high schools of New York: Boys' High School (Brooklyn), Wadleigh High School (New York City), De Witt Clinton High School (New York City). Three hundred ninety-eight first-year pupils were used in the experiment: 215 of Allen's original pupils, 80 girls from Wadleigh, and 103 boys from De Witt Clinton. Allen's pupils were those still remaining in school and available from the 11 Latin classes of the first semester of the year before. The Wadleigh and De Witt Clinton pupils were in three classes at each school. One class was selected at the upper, one at the middle, and one at the lower range of ability as determined by the Terman Group Test at Wadleigh, and the Otis at De Witt Clinton. The average age of all pupils was approximately 14 years.

The original analysis of the problem was made on the basis of "What are the factors which may possibly influence a pupil's success in first year Latin?" Some factors were eliminated because they appeared too subtle and elusive for our present scales of measurement; others because data could not be secured, or if at all, only with too great difficulty; others because they would not lend themselves to statistical treatment. Of the factors retained, it was not presumed at the outset that each had equal reliability when taken at its face value. One of the important aims of this study was to find to what degree the various factors were reliable for different groups.

Criterion

The criterion used in the experiment was a group of Latin Tests given at the end of the semester. For Boys' High School the scores of Dr. Allen were used, that is, the scores on the 11 Latin tests devised by Professor Briggs. One test was given to each of the following fields: Syllabification, gender, nouns, pronouns, conjugation, vocabulary, derivation, construction, pronunciation, translation from Latin

to English, translation from English to Latin. For the Wadleigh and De Witt Clinton groups, the pronoun test was omitted inasmuch as the topic had not been studied. Otherwise, the ten tests were made by the writer to parallel those of Professor Briggs. The tests were based on the materials covered in the textbook used in each individual school. Each test was so constructed that it could be scored in objective units, so that the poorest pupil could make some score, and so difficult that the best pupil could not make a perfect score. The methods of scoring, timing, and administration, were similar to those of any good standardized test.

Method of Treatment and Results

(*A*) *Raw Correlations.*—Each of the criterion tests was given a gross weighting, which was the median judgment of 16 Latin teachers and specialists. A combined weighted criterion score which took

Factor I: Allen Prognosis Battery of Six Tests

	Boys	Wadleigh	De Witt Clinton
Briggs Analogy, Alpha	.45	.43	.52
Birggs Analogy, Beta	.53	.50	.47
Thorndike Test of Word Knowledge, A	.35	.30	.51
Thorndike Test of Word Knowledge, B	.34	.18	.48
Rogers' Interpolation, 1	.26	.08	.41
Rogers' Interpolation, 2	.14	.34	.42

account of the standard deviation of each test was computed for each pupil. The combined weighted criterion scores and the scores on all factors were then transmuted, and coefficients of correlation of all factors computed by the Toops' method. Many details of the method of procedure and results found cannot be taken up in a discussion of this length.[1]

Approximately 60 items were considered for each of the three groups. They were classified under 16 heads called factors. Factors V, VI, and VIII were omitted from the Boys' High group. Below

[1] For complete procedure see: Clem, Orlie M.: Latin Prognosis, A Study of Detailed Factors of Individual Pupils. *Teachers College Contributions to Education*, Published by Bureau of Publications, Teachers College.

are summarized the raw coefficients of correlation of the various factors with the criterion. A note is given as to method where necessary, and also interpretative comments. The correlations are given in parenthesis. The first one given is for Boys' High School, the second for Wadleigh, and the third for De Witt Clinton.

Of all the objective factors used in the experiment, Briggs Analogy Tests, either Alpha or Beta, offer the best single basis of prognosis. The correlations for these tests approximate .50. Thorndike Tests of Word Knowledge have a much lower correlation and vary a great deal more. The correlations for the Rogers' Interpolation Tests are much lower and are very erratic for the three groups.

FACTOR II: INTELLIGENCE QUOTIENT (.36; .48; .61)

For groups (1) and (3) the Otis Group Test of mental ability was used, and for group (2) the Terman. It is not presumed that the IQ's are as accurate for these tests as for the Stanford Revision of the Binet-Simon. The above correlations are about what we should normally expect. Correlations of IQ with success in academic subjects have usually run from .40 to .60.

FACTOR III: AGE (−.27; −.38; −.57)

Age, seldom, if ever, in grade groups give other than a negative correlation with academic success. The brighter pupils get into high school earlier because they are bright.

FACTOR IV: HIGH SCHOOL ATTENDANCE (−.03; .27; .15)

FACTOR V: ELEMENTARY ATTENDANCE FOR LAST YEAR (——; 0.8; .07)

Both high school and elementary school attendance appear relatively unimportant factors in success in first year Latin. Although this is empirically true, few doubt the effect of long periods of absence. Long periods of absence usually mean elimination and hence no criterion measures are secured for these pupils.

FACTOR VI: ELEMENTARY SCHOOL MARKS FOR THE LAST YEAR

Elementary school marks for the last year have a very important value as instruments of prognosis in first-year Latin. The combined average of all marks (——; .41; .34) is a better predictor than a mark in a separate subject. Of the single subjects, arithmetic and English should be given precedence. Penmanship, music, and drawing have the lowest correlations.

Factor VII: High School Marks for the Semester, Except Latin

These rank by subject in practically the same order as corresponding subjects of the elementary school. The correlations of high school marks, due to recency, are much higher. Mathematics and English rank at the top; and drawing, music, penmanship, physical training, at the bottom.

Factor VIII: Teacher's Ranking of Pupils on Traits

Each Latin teacher ranked his or her pupils on the following twelve traits: Perseverance, industry, earnestness, nerve stability, orderliness, self-confidence, accuracy, right attitude toward criticism, frequency in securing help from teacher, promptness and regularity in doing work, ability to work independently, desirable social and moral attitudes. In ranking, each teacher was asked to give those in the upper 10 per cent a rank of (1), those in the next 20 per cent a rank of (2), the next 40 per cent a rank of (3), the next 20 per cent a rank of (4), the next 10 per cent a rank of (5). The correlations run from .32 to .67. All of them are consistently high. "Frequency of help" is the lowest. The closeness of the correlations would seem to indicate that the teachers did not distinguish carefully between the various traits; that is, if a pupil were given, for example (5) on industry, the tendency was to give him (5) on all others. We should expect, however, a reasonably high correlation between teachers' judgments of desirable traits.

Factor IX: Study and Conditions for Study

The correlations of the amount of outside help on Latin (.01; −.04; −4.09) seem to indicate that help is a negligible factor. Not only are the correlations negligible, but the pupils actually receive little help. These negative correlations are not startling for the reason that the duller a pupil is, the more help he will probably need.

The correlations for the average number of movies attended per week (−.09; −.17;−.17) are negative in the three groups. This would seem to justify in part the common assumption that the more movies pupils attend, the more poorly they do in their work.

The average number of hours of daily sleep (.007; −.06;.12) appears to be of little consequence.

Factor X: Individual Interests and Ambitions

The correlations for academic interest as shown by "plan after graduation" (−.04; .20; .18) and as shown by "plan for life work" (−.17; .12; .03) are not significant. In each case there is too small distribution because practically all who were taking Latin intended to go to college.

Factor XI: Outside Work

The correlations for time spent on music lessons are slightly but not significantly positive (.03; .18; .09). This may mean that the musical people are the brighter, or that they are the more ambitious. The same applies equally well to time spent on the study of any language or any school subject outside of school. The correlations are (.10; .12; .10).

The correlations for time spent in work for parents (−.04; −.005; −.10) are negative, but virtually negligible.

Factor XII: Amount of Home Study

Most of the correlations for the time spent in study of various high school subjects are negative. This is naturally what we should expect in subjects other than Latin, that as pupils spend more time in other subjects they do poorer in Latin. But the correlations for Latin itself are consistently negative (−.15; −.14; −.19). Hence, we may reasonably infer that the pupils who did poor work in Latin had to study more because they were dull.

Factor XIII: Ranking of Latin by Pupil on Basis of How Important He Considers It to Him

Each pupil ranked all subjects on the basis of how important he considered them to himself. The correlations are not particularly significant (−.10; .16; .14).

Factor XIV: Ranking of Latin by Pupils on Basis of Preference

The plan of ranking was the same as the above. The correlations are considerably positive (.27; .14; .52).

Factor XV: Ranking of Teacher by Pupil on Basis of Preference

The Latin pupils of group (3) were all taught by one teacher, and the high correlations of preference for subject (.52) and preference

for teacher (.38) indicate that both the subject and the teacher were popular with the pupils.

The correlations of preference for teacher of the Boys' High School and Wadleigh groups would indicate that preference for teacher has negligible effect (.10; −.04; .38).

FACTOR XVI: PARTICIPATION IN EXTRA-CURRICULAR ACTIVITIES

The correlations for the number of activities engaged in, fluctuate closely around zero (.01; .08; −.04).

(*B*) *Multiple Correlation, and the Selection of Prognosis Factors.*—The main problem of this study was to find the influence of certain detailed factors upon a pupil's success in first-year Latin, to choose the most effective factors, and through multiple correlation to obtain their combined effect as a basis for prognosis. "A multiple correlation coefficient is that correlation which expresses the total efficiency of the scale when tests chosen are those that bear the best or the highest correlation with the criterion."[1]

The contribution which any factor makes to a multiple correlation coefficient depends upon two considerations:

1. The correlation of the particular factor with the criterion.
2. The inter-correlation of the particular factor with other variables used.

Hence, to the investigator in prognosis, the desiderata of a good test or an effective factor are:

1. That it should correlate highly with the criterion.
2. That it should have a low correlation with other factors.

The formula used for the multiple ratio correlation coefficients in this experiment is that of Dr. Herbert Anderson Toops:

$$r_1c' = \sqrt{\frac{r^2_{1c} + r^2_{1\mu}{}' - 2r_{1c} \cdot r_{1\mu}{}' \cdot r\mu' c}{1 - r^2\mu' c}}$$

The formula gives correlation coefficients slightly less than the true multiple, but it is a very close approximation to it. It is an algebraic transmutation of some of the older formulæ, designed to economize time. It is not necessary to compute all the intercorrelations at the beginning of the process.

The importance of any factor among a series of factors depends upon its comparative ability to function when placed in the above

[1] Burr, Emily Thorp: Psychological Tests Applied to Factory Worker. *Doctor's Dissertation*, Columbia University, May, 1922, p. 73.

formula; that is, upon its ability to raise the coefficient of correlation. In general, the plan is to choose as a basic test or factor that one which has the highest correlation with the criterion, then to test the remaining factors for their contributions, to add the one which contributes most, and so on.

By the method of procedure explained briefly above, the data of the 60 items of this experiment were sifted for items in common with success in first-year Latin. For the three groups the four most objective and most consistently outstanding factors were: Briggs' Beta, age; Thorndike B; Elementary school average for the last year. The table following shows the contributions made by each of the three groups. Elementary average was lacking for group (I).

	Boys High	Wadleigh	De Witt
Briggs Beta, alone	.53	.50	.47
Briggs Beta, age	.55	.61	.66
Briggs Beta, age, Thorndike B	.56	.62	.71
Briggs Beta, age, Thorndike B and elementary average	...	.65	.72

By the further application of the multiple ratio correlation formula, the following prognosis batteries were selected:

	Boys High	Wadleigh	De Witt
1. Four Basic Factor Combination: Briggs Beta, age, Thorndike B, elementary school average.	.56[1]	.65	.72
2. Six Factor Combination: Briggs Beta, age; Thorndike B, elementary average, sum of traits (H. S. English and Math. average).	.66[2]	.82	.84
3. Allen Prognosis Battery: Briggs' analogy, Alpha and Beta; Thorndike Word Knowledge, A and B. Rogers' Interpolation, 1 and 2.	.56	.56	.67
4. Allen Prognosis Battery, plus Age and Elementary Average:	....	.66	.75

[1] Elementary average lacking.

[2] Elementary average and sum of traits lacking.

The first combination consists of four predictive factors, that is, factors for which objective data may be secured before the pupil begins the study of Latin. The second combination consists of four predictive factors, a fifth which is "affective," and a sixth which is predictive in the event that first-year Latin be not commenced until the second semester or later. This combination shows the achievement of pupils of a given age and capacity who in the judgment of the individual teacher possess certain traits of character and industry. The third combination is predictive and consists of six objective factors. The fourth combination consists of eight objective factors all of which are predictive.

It will be noted that the correlations for Boys' High are lower than those for Wadleigh and De Witt. Two reasons for this are evident:

1. The group has been made more homogeneous by the selection of 215 of Allen's original pupils.

2. Some factors are lacking for this group. Otherwise, as indicated throughout the experiment, the correlations of the Boys' High group would approximate those of the other two.

The correlations for the six factor combination for Wadleigh and De Witt are respectively .82 and .84. The mean standings of the pupils are as follows:

	Wadleigh	DeWitt
Six Factor Combination	10.19	8.33
Latin Criterion	11.8	11.1

In relation to pupil achievement, as measured by teachers' marks at the end of the semester, we find the following:

Wadleigh

Of the 80 pupils, 11 or 13.75 per cent failed Latin for the semester. Although the criterion was not used as final examination, 100 per cent of the failures were below the average on the criterion. If we choose some arbitrary score on the six factor combination such as (3), we find that 18.2 per cent of the failures and no one who passed fall below it; 27.3 per cent of the failures and no one who passed fall below (4); 72.7 per cent of the failures fall below (5), that is, in the lowest decile, with no one who passed, included.

De Witt Clinton

In the case of the Six Factor Combination with the De Witt Clinton group, 40 per cent of the failures, and one pupil who passed, are below (3). The mark of this one pupil is 65 per cent.

Of the failures 90 per cent and six pupils who passed are below (4). The average of these four pupils is 66.66 per cent with no one above 75 per cent; 100 per cent of the failures and 10 who passed are below (5), that is, in the lowest quartile. These 10 have an average mark of 69, with no one above 75. The lowest decile includes 60 per cent of the failures with only two who passed. The average mark of these two is 70.

The above analysis shows that the Six Factor Combination has a very high selective efficiency. Of the failures at Wadleigh, 72 per cent with no one who passed included, scored in the lowest decile, by the above prognosis method. At De Witt Clinton, 60 per cent of the failures with only two who passed, scored in the lowest decile.

FORMULAS FOR THE DETERMINATION OF CERTAIN MEASURES OF CENTRAL TENDENCY AND VARIABILITY FOR A COMPOSITE GROUP

FRANK C. TOUTON

University of Southern California

Problem.—To set down formulas for the computation of measures of central tendency and variability for a composite group using known measures of frequency of central tendency, and of variability for each of two or more related groups.

It frequently happens that one requires for statistical purposes a measure of central tendency and of variability for a composite group, which is made up of two or more groups for each of which such measures and the number of cases involved are known. The writer has recently found it necessary to treat as one group, and hence to determine measures of central tendency and variability, college entrants who had taken intelligence examinations at different dates and had been classified by semester groups, by sex, and by plan of admission used. For each of these sub-groups the frequency, the mean, and the standard deviation had been determined and recorded.

To meet this situation use was made of the formulas here stated.

(*A*) *Formula for determining the mean of the composite group:*

$$M_t = \frac{N_a M_a + N_b M_b + N_c M_c + \ . \ . \ .}{N_t},$$

in which

N_a, N_b, N_c, etc., represent the number of cases in groups *A*, *B*, *C*, etc.,

$N_t = N_a + N_b + N_c + \ . \ . \ .$ = the number in the composite group,

M_a, M_b, M_c, etc., represent the means of the several groups, and M_t = mean score of the composite group.

The above is a well known and commonly used formula, and is here quoted because of the economy of time which it makes possible.

(*B*) *Formula for determining the standard deviation of the composite group:*

$$\sigma_t = \sqrt{\frac{N_a(\sigma_a^2 + M_a^2) + N_b(\sigma_b^2 + M_b^2) + N_c(\sigma_c^2 + M_c^2) \ . \ . \ .}{N_t} - M_t^2},$$

in which

σ_t = the standard deviation of the composite group,

N_a, N_b, N_c, etc., represent the number of cases in the several groups,

$N_t = N_a + N_b + N_c + \ldots$,

M_a, M_b, M_c, etc., represent the mean scores of the several groups,

M_t = the mean score for the total group, and

σ_a, σ_b, σ_c, etc., represent the standard deviations of the several groups.

The above formula reduces to

$\sigma_t = \sqrt{\frac{N_a(\sigma_a^2 + M_a^2) + 0}{N_a} - M_a^2} = \sqrt{\frac{\Sigma(S_a^2)}{N_a} - M_a^2}$ for a single series and therefore holds for a single Series (A) with $N_a = N_t$, and $M_a = M_t$ and since $\Sigma(S_a^2)$ = the sum of the squares of the raw scores for Series (A), the formula reduces to the raw-score formula for the standard deviation of the scores in a single series.

This raw-score formula, $\sigma = \sqrt{\frac{\Sigma S^2}{N} - M^2}$, for the standard deviation of a series of scores can be readily deduced from the more commonly used deviation formula for the computation of the standard deviation of a series of scores, $\sigma = \sqrt{\frac{f \times d^2}{N}}$, as is shown in the content of the note given below.

The above formula can be expanded as follows:

$\sigma = \sqrt{\frac{f \times d^2}{N}} = \sqrt{\frac{\Sigma(d^2)}{N}}$ (where the variable d stands for the deviation of each of the several raw scores in turn from the mean and $\Sigma(d^2)$ indicates the sum of the squares of the several deviations).

$\sigma = \sqrt{\frac{\Sigma(Score - Mean)^2}{N}} = \sqrt{\frac{\Sigma(S - M)^2}{N}}$ (where the variable S stands for each of the several raw scores, in turn, and M is the mean of the series of scores).

Then expanding:

$$\sigma = \sqrt{\frac{\Sigma(S^2 - 2SM + M^2)}{N}}$$

$$= \sqrt{\frac{\Sigma S^2 - \Sigma(2SM) + \Sigma M^2}{N}}$$

$$= \sqrt{\frac{\Sigma S^2 - 2M\Sigma(S) + NM^2}{N}}$$

$$= \sqrt{\frac{\Sigma S^2 - 2M(NM) + NM^2}{N}}$$

$$= \sqrt{\frac{\Sigma S^2 - 2NM^2 + NM^2}{N}}$$

$$= \sqrt{\frac{\Sigma S^2 - NM^2}{N}}$$

$$= \sqrt{\frac{\Sigma S^2}{N} - M^2}$$

$$= \sqrt{\frac{\text{Sum of the squares of the several scores}}{\text{Number of the scores}} - (\text{Mean})^2}$$

$$= \sqrt{\text{Mean of the squares} - \text{Square of the mean.}}$$

The foregoing formula for the standard deviation of a single series of scores, in terms of the raw scores and their mean, is suggested by Franzen in the April, 1924, number of the *Journal of Educational Psychology* being derived by him by a similar method and here quoted because of its use in this article. The worker in the field of statistics will find that much time is saved in computing the standard deviation of a series of scores through the use of this formula along with a table of square and square roots and the Monroe Calculator or any standard adding machine.

Since $$\sigma = \sqrt{\frac{\Sigma S^2}{N} - M^2}$$

then $$\sigma^2 = \frac{\Sigma S^2}{N} - M^2,$$

and $$\sigma^2 + M^2 = \frac{\Sigma S^2}{N},$$

and $$N(\sigma^2 + M^2) = \Sigma S^2. \qquad (1)$$

Now since the variable S can represent the several scores in any series, it follows directly from the general formula, $\sigma = \sqrt{\frac{\Sigma S^2}{N} - M^2}$, that the constant ΣS_t could represent the sum of the several scores in a composite series made up of two or more related series, with N_t then representing the total number of cases in the composite group, and M_t the mean of the composite group.

Then from $\sum S_t^2 = \sum S_a^2 + \sum S_b^2 + \sum S_c^2 + \ . \ . \ . \ ,$

and

$$\sigma_t = \sqrt{\frac{\Sigma(S_t^2)}{N_t} - M_t^2},$$

we have, $$\sigma_t = \sqrt{\frac{\Sigma S_a^2 + \Sigma S_b^2 + \Sigma S_c^2 + \ \cdot\ \cdot\ \cdot}{N_t} - M_t^2} \qquad (2)$$

Since for the variable S, we have in (1) above $\Sigma(S^2) = N(\sigma^2 + M^2)$, then by substituting from (1) this value of ΣS^2 in (2), the above value of σ_t, we have,

$$\sigma_t = \sqrt{\frac{N_a(\sigma_a^2 + M_a^2) + N_b(\sigma_b^2 + M_b^2) + N_c(\sigma_c^2 + M_c^2) + \ \cdot\ \cdot\ \cdot}{N_t} - M_t^2}$$

This formula (2), provides a means of computing the standard deviation of a composite group from the frequencies, the means, and the standard deviations of the several related groups which enter into the composite group.

Use of Formulas Illustrated.—An application and verification of the formulas listed above is made in the computations of the mean and standard deviation of a composite group from data on the number of cases, the means, and the standard deviations for each of three distributions quoted from McCall's "How to Measure in Education," p. 407.

Given the following frequency distributions:

Score interval	Mid-point score (S)	Frequency A	Frequency B	Frequency C	Frequency (A + B + C)	Frequency × S	Frequency $\times S^2$
0– 2	1	3	1	1	5	5	5
2– 4	3	4	1	1	6	18	54
4– 6	5	4	2	1	7	35	175
6– 8	7	5	2	0	7	49	343
8–10	9	5	4	1	10	90	810
10–12	11	6	5	4	15	165	1815
12–14	13	4	4	2	10	130	1690
14–16	15	4	3	0	7	105	1575
16–18	17	3	2	0	5	85	1445
18–20	19	2	1	2	5	95	1805
		$N_a = 40$	$N_b = 25$	$N_c = 12$	$N_t = 77$	$\Sigma fS = 777$	$\Sigma fS = 9717$
Mean.....(m).....		9.55	10.76	10.50	10.091(x)		
Standard Deviation		5.08	4.39	5.30	4.94(y)	$\left(\text{using SD} = \sqrt{\frac{\Sigma S^2}{} - M^2}\right)$	

1. *To show that:*

$$\text{Mean}_{(A+B+C)} = \frac{N_a M_a + N_b M_b + N_c M_c}{N_t}$$

$$= \frac{40 \times 9.55 + 25 \times 10.76 + 12 \times 10.50}{77}$$

$$= \frac{777}{77}$$

$$10.091 = 10.091 \qquad \text{(From } (x) \text{ above)}$$

2. *To show that:*

$$SD_t = SD_{a+b+c} =$$

$$\sqrt{\frac{N_a(SD_a^2 + M_a^2) + N_b(SD_b^2 + M_b^2) + N_c(SD_c^2 + M_c^2)}{N_a + N_b + N_c} - M_t^2}$$

$$\sqrt{\frac{9717}{77} - 10.091^2} =$$

$$\sqrt{\frac{40(5.08^2 + 9.55^2) + 25(4.39^2 + 10.76^2) + 12(5.30^2 + 10.50^2)}{77} - 10.091^2}$$

$$\sqrt{126.194 - 101.828281} =$$

$$\sqrt{\frac{4680.36 + 3376.25 + 1660.08}{77} - 101.7617}$$

$$\sqrt{24.365719} = \sqrt{126.191 - 101.7617}$$

$$4.94 = \sqrt{24.4293} \text{ or } 4.94. \qquad \text{(From } (y) \text{ above)}$$

SPELLING ABILITY AND MEANING VOCABULARY AS INDICATIONS OF OTHER ABILITIES

JOHN A. LESTER

Pottstown, Pa.

The object of this study is to determine to what extent a knowledge of the spelling ability and the reading vocabulary of a student might serve as an indication of his ability to do school work required in the last two years of the college preparatory course. Practically all of the students examined went or will go to college. The purpose was to facilitate the practical adjustment of intellectual work to intellectual power under specific conditions; and though the results are useful only as interpreted in the light of those conditions, they are worth recording.

Spelling Ability

Assuming that grades which record the accuracy of work done in column spelling drill may be accepted as a measure of spelling ability, we may calculate the coefficient of correlation between this ability, and other abilities indicated by grades. The results of investigations of this nature may first be mentioned. Houser[1] found in the high fourth, high sixth, high seventh, and high eighth grades in the Haight School in Alameda, California, that the correlation between general intelligence and spelling ability as determined by column spelling was +0.5304. Murray[2] found that in Sweet Briar College, Virginia, the correlation between academic grades and spelling ability as represented by tests in column spelling was +0.42.

The correlations which I give below in Table I were calculated on the same basis as those of the two investigations described. The work

Table I.—Correlation of Spelling Ability as Determined by Column Spelling Drill Grades, with Ability in Other Subjects as Determined by Grades

	French	Latin	Geometry	Algebra	School standing
Correlation.....	.39	.36	.33	.28	.55

[1] *Elementary School Journal*, Vol. XVI, pp. 190–199.

[2] *Journal of Educational Psychology*, Vol. X, pp. 357–376.

of 98 students in the same material was examined in two groups of 52 and 46. The school grade was reckoned uniformly in every subject, the average of grades for daily work counting 50 per cent, and the grade for the final examination upon that work counting 50 per cent. The school standing was uniformly determined from the grades in the different subjects by averaging these grades, and weighting each according to the time expended on it. It is to be noted that spelling ability here is measured by accuracy in writing in column, from dictation, words previously studied.

It is however clearly illogical to regard a correlation calculated on the basis of that just described, as correlation of spelling ability with this or that other ability. The ultimate test of spelling ability is the degree of accuracy attained in the writing of words in actual spontaneous composition. While drill in column spelling is the readiest and most useful means of acquiring that ability it is not in itself its proper test or measure. The words used in column drill may have no relation to the student's actual vocabulary in spontaneous writing. Indeed, Houser took for the basis of his investigation every hundredth word from an unabridged dictionary. But some 25 per cent of the commonest mistakes in spelling are mistakes in derived words which do not occur in the dictionary; and of the total mass of misspellings observed more than 50 per cent are lapses or errors of inattention.[1] But when pupils write a column of 10, 15, or 20 single and separate words and have time to review them, this kind of liability to error is materially reduced. Hence column spelling by its very nature measures imperfectly the most serious contributing cause of bad spelling as it appears in the ordinary communication of life. Finally column spelling takes no cognizance of the frequent mistakes in word-compounding and in the use of the apostrophe to denote possession—two kinds of error which together account for more than 24 per cent of all the errors made in spontaneous composition.

Attempts to calculate the coefficient of spelling ability as based upon accuracy in spontaneous composition, involving as they do accurate records of individual misspellings over a considerable extent of written work, have been infrequent. Brandenburg[2] found a correlation between spelling ability in composition and general scholarship in certain classes at the University of Wisconsin which is represented by misspelling averages of 1.8, 3.8, and 6.5 for scholarship grades of 90

[1] See *Journal of Educational Psychology*, Vol. XIII, p. 73.

[2] *School and Society*, Vol. VII, p. 26.

per cent or over, 80–89 per cent, and below 80 per cent respectively. F. M. Foster[1] concludes from his investigation of the spelling of Freshmen at the University of Iowa that correlation between mental ability and spelling ability is "quite remarkable."

For many years in the teaching of English at the hill school, Pottstown, Pa., I have required my students to record with scrupulous care in convenient forms provided, every word they misspell in their weekly compositions. At the end of the school year each student arranged his individual list of words alphabetically, indicating their frequencies, the classes of error into which they fell, and their sum total. These records afforded material for the classification of students in accordance with actual spelling ability—that is, the ability to write in spontaneous composition the correct forms of the words they had occasion to use. From these records the correlations of spelling ability with other abilities as determined by school grades were calculated. The method used was the Pearson method of rank differences.

The annual group consisted of 60 to 70 boys of the average age of 17. The correlations for Latin, French, and geometry represent groups slightly smaller, the correlation for history a group of about 35. In Table II is shown the record of these correlations with the mean correlation for the period of nine years. It should be stated

TABLE II.—CORRELATION OF SPELLING ABILITY AS DETERMINED BY ACCURACY OF SPELLING IN SPONTANEOUS COMPOSITIONS, WITH OTHER ABILITIES AS DETERMINED BY SCHOOL GRADES

Year	School standing	English	Latin	French	Geometry	History
1915–16	.6395	.5211	.5227	.3086	.2580	.4801
1916–17	.5996	.4929	.4120	.3571	.2901	.3531
1917–18	.6386	.5315	.4371	.3917	.3075	.3952
1918–19	.5700	.4920	.3915	.3823	.3501	.3882
1919–20	.6600	.5579	.4602	.3914	.4328	.5104
1920–21	.6191	.4895	.4234	.3571	.3261	.4462
1921–22	.5971	.5281	.4678	.3465	.3746	.4327
1922–23	.6248	.4932	.5572	.5334	.3038	.3556
1923 autumn term	.5962	.5058	.4956	.4155	.3429	.2851
Mean.....	.6161	.5124	.4631	.3871	.3317	.4052

[1] *School and Society*, Vol. V, pp. 506–8.

that the uncertainty of correlations of this kind which arises from variability of teachers' estimates is minimized by the fact that the administration of each of the departments concerned has been in the same hand from 1915 to 1923, and hence the standard of judgment in any given department has been uniform.

It should be noticed that in the second column of correlations we are in a degree correlating spelling ability with itelf, in so far as correct spelling was one of the contributing factors in the English grades. But this objection is not applicable to any of the other correlations, except again insofar as the English grade is a part of the composite grade which established the school standing of any particular student.

Reading Vocabulary

For the past four years I have kept the records of vocabulary tests which I have used in the study of these groups of pupils. The Whipple Test was used every Autumn, and in addition the test recently made available by Professor Inglis, of Harvard University, was given in November of 1923. The annual group, as in the investigation of spelling ability, consisted of 60 or 70 boys of 17 years of age, the personnel being the same as for the last four years recorded in Table II. The correlations of reading vocabulary with other abilities as determined by school grades is given in Table III. For convenience of comparison both the Whipple and the Inglis correlations are included under 1923, but the mean for the four years is calculated from the Whipple correlations alone.

Table III.—Correlations of Reading Vocabulary as Determined by the Whipple and Inglis Tests with Other Abilities as Determined by School Grades

Year	School standing	English	Latin	French	Geometry	History
1920	.5125	.5543	.2560	.2325	.2560	.6180
1921	.3124	.5931	.3108	.2305	.3271	.6574
1922	.3866	.5844	.3279	.2178	.3094	.6378
Whipple	.3124	.5931	.3107	.2305	.3271	.6577
Inglis (1923)	.4425	.6086	.3216	.2143	.3513	.6855
Mean	.3810	.5812	.3014	.2278	.3049	.6427

The mean correlations of Tables II and III are brought together for facility of comparison in Table IV, with the Inglis correlations added in parentheses.

TABLE IV.—COMPARATIVE MEAN CORRELATIONS OF SPELLING ABILITY AND VOCABULARY WITH OTHER ABILITIES AS DETERMINED BY SCHOOL GRADES, REPEATED FROM TABLES II AND III

	School standing	English	Latin	French	Geometry	History
Spelling ability.....	.6161	.5124	.4631	.3871	.3317	.4052
Meaning vocabulary.....	.3810	.5812	.3014	.2278	.3049	.6427
Inglis.....	(.4425)	(.6086)	(.3216)	(.2143)	(.3513)	(.6855)

INTERPRETATION OF THE CORRELATIONS

It should be noticed that correlations of other abilities with reading vocabulary as measured by the Inglis Test are almost uniformly higher than the similar correlations as calculated from the results of the Whipple Test. This fact would indicate that a perfect measure of meaning vocabulary would show higher correlations than the mean correlations of Table III; for the Inglis Test is in the first place more widely based, containing 150 words as against 100 in the Whipple Test; and in the second place by its very nature it eliminates the danger of overestimation or underestimation of his vocabulary by the student—a tendency which at best can be only partially safeguarded by the investigator using the Whipple Test. The results of the two tests taken by the same group of 56 boys show a correlation of .8755.

The measure of spelling ability used in this investigation indicates the degree of automatic accuracy acquired by the pupil in writing the forms of words when his main attention is fixed on something else. The correlation between this ability and other abilities as measured by school grades is higher than the corresponding correlations in all subjects except English and history, and this in spite of the fact that spelling ability is a factor in the English grade correlated. The other vocabulary correlations are almost negligible.

These high correlations of the results of vocabulary tests with ability in English and history are in accordance with *a priori* judgment

that the student who can read with ease will get the best results in subjects in which that ability counts for most. Such correlations have proved useful in specific cases in the guiding of students towards those studies for which they seemed best fitted.

It is to be noted that the correlation of spelling ability with other abilities is highest for the composite of those abilities labeled as "school standing." The frequency and nature of errors of spelling in spontaneous composition, often apparently insignificant and trivial in themselves, have a real value as a measure of that accuracy which is essential to other pursuits and studies; and particularly as revealing the habits of mind adapted for college work. Houser's conclusions in the investigation referred to, that there is a closer correlation between spelling ability and "general intelligence" than between meaning vocabulary and "general intelligence"—"general intelligence" being measured by teachers' estimates of it, is borne out by these results.

Influence of the Study of Latin on Reading Vocabulary and Spelling Ability

What follows is apart from the main object of this study. It relates to the influence of the study of Latin on the enlargement of the student's reading vocabulary in English, and upon the accuracy with which he writes that vocabulary in spontaneous composition.

By dictating lists of words to students in the University of Iowa, Foster[1] attempted to discover whether the high school Latin course had been of any value in giving ability to write Latin derivatives. He found no evidence that the study of Latin had been of material aid in that direction. Of the 56 students of the average age of 17 who wrote the Whipple and Inglis Vocabulary Tests last Autumn, some had studied no Latin, some had studied Latin for two years, and some had studied it for three years. I wished to find out what influence if any, this study of Latin had had on their reading vocabulary, and also upon their accuracy in spelling in spontaneous composition. This information is arranged in tabular form in Tables V and VI.

It would therefore appear that Latin as it was taught to these boys in their college preparatory courses had no effect either on their English vocabulary or upon the accuracy with which they wrote English words in their compositions.

---

[1] *School and Society*, Vol. V, pp. 506–8.

TABLE V.—INFLUENCE OF THE STUDY OF LATIN ON READING VOCABULARY, IN TERMS OF PERCENTAGE SCORES OF 57 BOYS OF THE AVERAGE AGE OF 17 IN THE WHIPPLE AND THE INGLIS VOCABULARY TESTS

		Reading vocabulary of the whole group of 57 boys	Reading vocabulary of those who had studied Latin for three years	Reading vocabulary of those who had studied Latin for two years	Reading vocabulary of those who had studied no Latin
Whipple	Median.....	67.9	68.7	68.1	67.2
	Median.....	67.0	70.0	66.0	66.0
Inglis	Median.....	66.4	66.2	69.7	64.0
	Median.....	66.0	66.0	67.0	66.0

TABLE VI.—INFLUENCE OF THE STUDY OF LATIN ON SPELLING ABILITY AS DETERMINED BY ACCURACY IN SPONTANEOUS COMPOSITION SPELLING ABILITY INDICATED BY PERCENTAGE SCORES

	Whole group of 57 boys	Latin for three years	Latin for two years	No Latin
Mean.....	65.8	64.6	67.5	67.1
Median.....	66.0	64.0	67.0	67.0

We may then summarize the results of this study in correlations in three statements. It seems to be true that under conditions prevailing in the preparatory school:

1. Accuracy in spelling in spontaneous composition is a valuable indication of a habit of attention which carries over into other studies.

2. That a high index in the vocabulary test is ground for expectation of success in preparatory school courses in English and history.

3. That the study of Latin under preparatory school conditions has apparently no effect on increasing the volume of English vocabulary nor accuracy in writing the forms of words in free composition.

A PROGNOSTIC TEST IN TYPEWRITING

MARY LYNCH GRONERT

Graduate Student in Psychology

Washington University, St. Louis

While teaching typewriting in the St. Louis high schools the writer was impressed with the need of a prognostic test for rapidity and accuracy in manipulating the keys, since many students after the usual period of preparation and probation found that they were unfit for the work. The following prognostic test invented by the writer was tested in two St. Louis high schools and its predictive value was found to be sufficiently high to warrant its continued use.

The Lynch Prognosis Test, as it is called, is a substitution test. The form of the test was determined after an investigation of the results of the tests given by Thorndike and others and described in Whipple's Manual of Mental and Physical Tests. Comparison of the results of the various methods of testing abilities indicated that a substitution test would best measure the mental traits necessary for progress in touch typewriting, namely, the ability to memorize quickly, mental alertness, and the ability to concentrate to such a degree that there shall be little, if any, interruption in the process which transforms the visual image of the letter upon the printed page into a motor response to strike that letter with the proper finger upon the typewriter.

Under the method of admitting to classes in touch typewriting all who applied, without first subjecting them to a prognostic test, the failures were well over 50 per cent. By having admission based on the score made in the Lynch Prognosis Test this percentage can be reduced to 17½, or less.

The test was given three times at intervals of six days and four days respectively, and a record of the number of correct substitutions made in each test by each pupil was made. In order to make the test as nearly an accurate measure as possible of the ability of the pupils, the time was made short enough so that no pupil could finish the entire test. The first test was given in one and one-half minutes; the second test was given in one and one-quarter minutes; and the third test was given in one minute. In each test the highest score made was 99 out of a possible 100 substitutions. Only one pupil made the highest record in each test.

The curtailing of the time for each test resulted in changing the median. In the first test the median was 62.5; in the second test 61.1, and in the third test 60.

Table A represents the tabulation of results from one class of pupils who were graded in typewriting by a teacher other than the one testing them. In Tables B and C the teacher who gave the substitution test graded the pupils in typewriting. The subjects in Table A being graded by a different teacher, represent a correlation in which the achievement of each pupil was unknown to the other teacher until the time for making comparisons. In this manner the personal element of the teachers was eliminated. The tabulations of results in Tables B and C represent the results from testing 130 pupils in the Cleveland (St. Louis) High School. The test was given to groups of about 30 pupils. Typewriting achievement is represented or measured in the following grades: F, for less than 65 per cent; P, for 65 to 69; M, for 70 to 79; G, for 80 to 89; E for 90 and above. The relation between typewriting achievement as measured by these grades and the score made in the substitution test is presented in Tables B and C.

In the key given below a number appears under each letter. When the instructor said "Go!" the subjects wrote below the letters in the test the number that appears under that letter in the key. The substitutions were made for the letters in the order in which they appeared in the test, that is, one line had to be completed in order and then the next line taken. There was no jumping allowed through the entire test.

KEY

A	M	Y	U	S	X	Z
1	2	3	4	5	6	7

U	M	Z	Y	A	S	M	X	A	U
X	S	M	Z	Y	X	A	U	X	Y
Y	Z	U	M	S	Y	X	A	S	M
S	Y	Z	U	X	A	S	M	Z	U
A	X	M	Z	Y	S	U	A	X	Y
Z	S	A	X	M	Y	S	U	A	Z
M	X	U	Y	S	A	Z	M	Y	A
Z	S	M	Y	U	Z	A	X	S	Z
A	U	X	M	S	Y	A	Z	U	X
U	M	Z	Y	X	A	S	Y	M	U

Name.....
Attempt #.....
Number right.....
Number wrong.....
Score.....

A preliminary comparison of typewriting achievement with the scores made in the substitution test seemed to indicate the following: First: Wherever the performance of the subject in the substitution test equalled or excelled the median for the class he was almost sure to do very good work in typewriting. Second: Wherever the performance of the subject was consistently 10 or more points below the median for the class he was almost sure to do failing work. Third: If the subject, though below the median for the class, showed ability to improve in the performance on the substitution test and his results approached the median for the class, he was regarded as being in the border line class and needed very careful watching and much encouragement over the learning plateaus. A line of demarcation was drawn at 50 on the three substitution scores and the tabulations were made accordingly.

TABLE A.—NUMBER AND PER CENT OF 28 PUPILS BELOW AND ABOVE LINE OF DEMARCATION (50) ON SUBSTITUTION SCORE, MAKING GRADES OF F, P, M, G, AND E, RESPECTIVELY, IN SUBSEQUENT TYPEWRITING WORK. FALL TERM 1921

Substitution score	F (Below 65)	P (65–69)	M (70–79)	G (80–89)	E (90–100)	Total
Below 50.....	2(40 %)	1(12½ %)	1(16⅔ %)	0(0 %)	0(0 %)	4(14¾ %)
50 and above.....	3(60 %)	7(87½ %)	5(83⅓ %)	6(100 %)	3(100 %)	24(85¾ %)
Totals.....	5	8	6	6	3	28

TABLE B.—NUMBER AND PER CENT OF 130 PUPILS BELOW AND ABOVE LINE OF DEMARCATION (50) ON SUBSTITUTION SCORE, MAKING GRADES OF F, P, M, G, AND E, RESPECTIVELY, IN SUBSEQUENT TYPEWRITING WORK. FALL TERM 1921

Substitution score	F (Below 65)	P (65–69)	M (70–79)	G (80–89)	E (90–100)	Total
Below 50.....	20(87 %)	10(62½ %)	6(28 %)	3(7 %)	0(0 %)	39(30 %)
50 and above.....	3(13 %)	6(37½ %)	15(72 %)	43(93 %)	24(100 %)	91(70 %)
Totals.....	23	16	21	46	24	130

In considering Table B it is evident that 87 per cent of those who failed in typewriting made a score below 50 in the substitution test. On the other hand 100 per cent of the E's went to those who scored

TABLE C.—NUMBER OF PUPILS MAKING GRADES OF F, P, M, G, AND E RESPECTIVELY, IN TYPEWRITING IN RELATION TO SCORES MADE ON SUBSTITUTION TEST FOR FALL TERM 1921

Substitution score	F (below 65)	P (65–69)	M (70–79)	G (80–89)	E (90–100)	Total
0–29	2	0	0	0	0	2
30–34	4	0	0	0	0	4
35–39	3	1	0	0	0	4
40–44	8	5	1	0	0	14
45–49	3	4	5	3	0	15
50–54	2	5	5	7	0	19
55–59	0	1	6	9	0	16
60–64	1	0	2	11	4	18
65–69	0	0	1	3	2	6
70–74	0	0	1	4	3	8
75–79	0	0	0	5	4	9
80–84	0	0	0	2	5	7
85–89	0	0	0	2	2	4
90–94	0	0	0	0	2	2
95–100	0	0	0	0	2	2
Totals.....	23	16	21	46	24	130

above 50 in their prognostic test. The percentage of F's is more than six times as great in the group below as in that above, while the high grades are 13 or more times as frequent in the group above 50.

There have been further investigations made in the use of the Lynch Prognosis Test and the conclusions reached have confirmed those disclosed in this study.[1]

[1] See *St. Louis Public School Messenger*, May, 1924.

A COMPARISON OF THE MENTAL ABILITY AND SCHOOL ACHIEVEMENT OF THE BRIGHT AND DULL PUPILS IN THE SIXTH GRADE OF A LARGE SCHOOL SYSTEM

M. J. VAN WAGENEN

University of Minnesota

The present study is an outgrowth of two earlier ones, both of which need to be described briefly to give the basis of this study. The first one consisted of the determination of mental age equivalent scores for educational scales in spelling, reading, the informational and thought phases of American history, the informational and thought phases of geography, the fundamental operations of arithmetic and arithmetic problems. In several instances the scales themselves had to be devised. On the basis of the results obtained by giving educational scales in the eight phases of school work just mentioned as well as mental tests to some 1200 pupils about equally scattered between the two sexes and throughout the sixth, seventh and eighth grades, the equivalent scores were determined for the middle range of mental ages. On the basis of the average scores for each mental age the regression equations were worked out for each educational scale. With the use of these regression equations the equivalent scores were extended in each direction. Such equivalent scores were then carefully checked up for discrepancies with available grade norms and also with the results obtained from a carefully selected class of mentally superior pupils who were being pushed along to something like their full capacity by a very capable teacher. The final results of this study have been embodied in a chart devised for finding mental quotients, educational ages, educational quotients, and achievement or accomplishment quotients directly from chronological ages and the mental test and educational scale scores. The two significant features of the chart for this study are first; that the achievement quotients are based on mental ages, and second; that the basis of determining achievement quotients for those with either low or high mental ages is independent of any tendencies that may characterize the quality of work of mentally retarded or of exceptional pupils only.

The second study out of which the present one developed can be more briefly described. It consisted of a survey of the mentality

and achievement of all the pupils completing the sixth grade in the public schools of St. Paul, Minnesota, last June. In this survey one of the same mental tests was employed as was used in the original derivation of the chart—the Terman group test of intelligence, examination A. For measuring the achievement of the pupils either the same or equivalent educational scales were employed as were used in the first study. These consisted of a spelling scale from an unpublished series, one of the Thorndike-McCall reading scales, recent revisions of my own information and thought scales in American history, one of the information and thought scales from the Posey-Van Wagenen series of geography scales, a modified form of the Woody arithmetic scales, and one of the Buckingham arithmetic problems scales, with a slightly different method of finding the pupils' scores. All the derived measures, such as the mental ages, mental quotients, and achievement quotients, were determined with the use of the chart mentioned above.

From the materials available in the St. Paul survey there have been selected for the present study the mental quotients, the educational scale scores, and the achievement quotients of all sixth grade pupils—52 boys and 49 girls—with mental ages below 11 years 3 months, and of all the sixth grade pupils—39 boys and 54 girls—with mental ages of 14 years 3 months and above. The materials selected provide for a comparison of both the achievement and of the quality of work of the pupils who were 15 months or more below the normal mental age of 12 years 6 months, with the achievement and quality of work of the pupils who were 21 months or more above the normal mental age for the grade.

It is striking perhaps that the mentally oldest pupil in the lower one-thirteenth of the group differs in mental age from the youngest pupil of the higher one-thirteenth by three years. The difference between the median mental ages for the two groups is 3 years 10 months for the girls and 4 years for the boys. The median mental quotient for the lower group of girls is 79, for the higher group 128, as contrasted with a median mental quotient of 101 for the entire group of sixth grade girls. For the boys, the median mental quotient is 81 for the lower group and 129 for the higher group as compared with 101 for the entire group of sixth grade boys.

While the normal pupil is making a year's growth mentally, the median pupil in the lower group is making a mental growth of about 9.6 months, and the median pupil in the upper group a mental

growth of 15.6 months. Assuming that the mental quotient is a reasonably reliable index of the rate of mental growth, it is evident that these pupils are becoming six months further apart mentally each year they are in school. Equally significant is the inquiry as to

TABLE I.—DIFFERENCES BETWEEN THE MEDIAN AGES AND MEDIAN EDUCATIONAL AGES OF THE TWO GROUPS OF PUPILS EXPRESSED IN YEARS AND MONTHS FOR ALL PHASES OF THE SCHOOL WORK MEASURED

Sex	Educational Ages in																	
	MA		Spelling		Reading		American history				Geography				Arithmetic			
							Information		Thought		Information		Thought		Fundamental		Problem	
	Year	Month	Year	Month	Year	Month	Year	Month	Year	Month	Year	Month	Year	Month	Year	Month	Year	Month
Girls	3	10	2	0	3	9	2	5	3	11	3	7	3	1	3	0	3	0
Boys	4	0	2	1	3	9	3	0	4	4	3	6	3	0	3	4	3	10

TABLE II.—DIFFERENCES BETWEEN THE MEDIAN EDUCATIONAL AGES AND CORRESPONDING MENTAL AGES FOR EACH GROUP OF PUPILS, AS WELL AS FOR THE ENTIRE GROUP OF SIXTH GRADE PUPILS, EXPRESSED IN MONTHS OF MENTAL AGE

Sex	Group	Spelling	Reading	American history		Geography		Arithmetic	
				Information	Thought	Information	Thought	Information	Thought
Girls	Lower	+9	−5	−35	−15	−20	−12	−1	−14
Boys	Lower	+8	−5	−34	−16	−16	−10	−4	−6
Girls	Entire	+1	−4	−40	−10	−20	−14	+2	−12
Boys	Entire	−4	−7	−37	−9	−13	−13	−1	−7
Girls	Higher	−13	−6	−52	−14	−23	−21	−11	−24
Boys	Higher	−15	−8	−46	−12	−22	−22	−12	−8

whether or not their educational ages develop in proportion to their mental ages. The differences in educational ages between the median pupils in each of these groups have also been computed for each phase

of school work tested and are given in the first table. For the purpose of interpreting the variations in the educational age differences a second table is given, showing the extent in months of mental age that the median scores for each group, as well as for the entire group of sixth grade pupils, are above or below the corresponding mental age norms.

A glance through the first table reveals the tendency for the educational age differences to be less than the mental age differences. The thought phase of American history is the single exception, and this holds for both the girls and the boys. In view of the fact that the probable errors of the differences are all less than 1.5 months, very few of the variations can be considered chance occurrences. Marked variations must be due either to an exceptional or at least a good quality of work on the part of the pupils with the lower mental ages, or else to a poor quality of work on the part of the bright pupils. Table II shows that in the very first case—spelling—the markedly smaller educational age differences are obviously due to both factors. The less capable pupils are apparently stimulated to learn to spell, probably beyond any need they will ever have for the ability, while less in the way of spelling ability in proportion to their mental capacity is expected of the brighter pupils. Evidently the attempt is being made to hold pupils of all mental ages to somewhat the same standard of spelling ability, irrespective of their social needs for their ability or their capacity to acquire it.

In the case of reading, where a comparison of the educational age differences and the mental age differences show but a small variation, the brighter pupils are doing a slightly superior quality of work to the duller pupils. The informational phase of history, however, shows a significant variation again, though not so marked a one as in the case of spelling. Table II demonstrates that this condition is not due to a superior quality of work on the part of the duller pupils, but that however poor the quality of work of the incapable pupils is, the work of the normal pupils is still poorer, while that of the bright pupils is even poorer yet. The thought phase of American history shows educational age differences wider than the mental age differences. While this condition is not due to a superior quality of work on the part of the bright pupils, it is clear from Table II that the bright pupils are doing a better quality of work than the dull pupils, even though it is poorer in quality than that of the pupils of average intelligence. In the informational and thought phases of geography the best quality of

work is done by the average pupils in the first instance and by the less capable pupils in the second instance, a fact which accounts for the differences in the educational ages in the two phases of geography in Table I. For the brighter pupils the fundamental operations in arithmetic show a condition similar to that in spelling. Although the work in arithmetic would seem quite as important to the duller pupils as the ability to spell, it received nothing like the same attention as far as difficulty of tasks is concerned. In the thought aspect of arithmetic there is a marked difference in quality of work among the different groups of girls, but a negligible one among the groups of boys, as indicated in Table II. This relatively poorer quality of work in the higher group of girls explains the narrower educational age difference for the girls in the last column of Table I.

A detailed analysis of the data presented in these two tables tends to make one more cautious in asserting that the best quality of work is invariably done by the less capable students. In these tables two significant tendencies stand out very clearly. With the exception of spelling alone more marked contrasts are evident from one phase of school work to another than from one mental age group of pupils to another. Quite true is the inference that this is undoubtedly due to the variation in emphasis and the adequacy of the methods of teaching. And equally true would probably be the inference that the contrasts between the mental age groups are due to the same circumstances.

More significant, however, are the sex contrasts. As one glances down the columns under spelling, reading, and the fundamental operations in arithmetic in the second table, he is struck by the fact that in every group the quality of work of the girls is better than that of the boys. Just the opposite tendency is evident, however, as one follows down the columns under the informational phases of American history and geography and under the thought phase of arithmetic. In the thought phases of both American history and geography the tendency is less pronounced, but in favor of the boys. That the sex variations in quality of work are quite as pronounced in the duller group as in the brighter group is also evident from the second table.

In view of the fact that the variations in months given in Table II bear no constant relation to chronological time, care must be taken not to interpret the data of Table II as if these were the numbers of chronological months the average pupils of the several groups would need to catch up with their mental age norms if their mental ages could be held still for the time being. Table III presents the same

data for the lower and higher groups in this form. The fact that in this form—the number of chronological months of retardation or acceleration—the data for the two groups are, on the average, more nearly the same, is suggestive that the mentally later school entrance of the brighter pupils as compared with that of the duller ones—at seven and a half years mentally as compared with five years—may be one of the factors accounting for the poorer quality of work of the brighter pupils, especially as it is so prominent even at the end of the sixth grade. That other factors, such as instructional neglect for instance, may be just as responsible or even more so is not to be doubted, but this one is more likely to be overlooked.

TABLE III.—DIFFERENCES BETWEEN THE MEDIAN EDUCATIONAL AGES AND THE CORRESPONDING MENTAL AGES FOR THE LOWER AND HIGHER GROUPS OF PUPILS WHEN REDUCED TO CHRONOLOGICAL VARIATIONS

Sex	Group	Spelling	Reading	American history		Geography		Arithmetic	
				Information	Thought	Information	Thought	Fundamental	Problem
Girls.....	Lower	+11¼	−6¼	−43¾	−18¾	−22½	− 9	− ¾	−10½
Boys.....	Lower	+10	−6¼	−42½	−17½	−17½	− 6	−1½	− 4½
Girls.....	Higher	− 9¾	−4½	−36	−12	−17¼	−15¾	−6¾	−18
Boys.....	Higher	−11¼	−6	−34½	− 9	−16½	−16½	−8	− 6

Although the form of presenting data used in Table II—retardation or acceleration in months from the mental age norms—brings certain characteristics out clearly, no attention is given to the effect of mental age. Should mentally late school entrance or absence from school be the only factor accounting for retardation, then the form made use of in Table II for presenting the data would be the most accurate one to use. As other factors, such as a lack of adequately rapid promotion to keep pace with the rate of mental growth, are likely to have a constant effect upon the quality of work a retardation of six months in achievement is of more significance at a mental age of 10 years than at a mental age of 15. The use of achievement quotients in Table IV for presenting practically the same data as is given in Table II eliminates this element and thus gives a more accurate indication of the effect of continuous conditions of instruction.

An examination of the materials in Table IV reveals the fact that in about half the cases the quality of work of the group as a whole is just as good or better than that of the mentally younger pupils. But, while the work of the brighter pupils shows just as good or an even better quality than that of the duller pupils in reading, in the thought phase of American history and in the informational phase of geography, in no instance does it show as good a quality as that of the entire group. This condition is suggestive that other factors that may account for the poorer quality of the brighter pupils' work are their mental retardation in the grades and either instructional neglect of the brighter pupils in mixed classes or furnishing them when segregated with an enriched curriculum of simple materials instead of materials sufficiently difficult and interesting to challenge their best efforts.

TABLE IV.—MEDIAN ACHIEVEMENT QUOTIENTS FOR EACH GROUP OF PUPILS AS WELL AS FOR THE ENTIRE GROUP OF SIXTH GRADE PUPILS

Sex	Group	Spelling	Reading	American history		Geography		Arithmetic	
				Information	Thought	Information	Thought	Information	Thought
Girls.....	Lower	108	96	74	88	82	90	103	87
Boys.....	Lower	106	95	77	88	90	93	98	95
Girls.....	Entire	103	98	72	92	86	90	102	90
Boys.....	Entire	99	96	75	94	90	91	98	93
Girls.....	Higher	93	98	70	91	85	87	91	83
Boys.....	Higher	91	95	74	93	87	87	92	91

STANDARD DEVIATIONS OF AGE SCORES AND QUOTIENTS IN TYPICAL GROUPS

G. M. WILLSON

Assistant Psychologist, New Jersey State Department of Institutions and Agencies
Prepared under the Auspices of the Division of Education and Classification[1]

Purpose

The chief purpose of this article is to indicate the need for a more definite knowledge of the standard deviations of mental ages, educational ages and intelligence quotients in typical age and grade groups and to suggest the problems which arise in the course of the collection of such sigmas. The actual sigmas reported are not sufficiently numerous to be conclusive.

Values of Data on the Standard Deviation of Typical Groups

1. Coefficients of reliability reported in the literature can be compared one with another only when the dispersions of the correlated measures in the several reported studies are known. Since a reliability coefficient of .80 in a group having a sigma of test age of 10 months is equivalent to a coefficient of .95 in a group having a sigma of test age of 20 months,[2] much of the data on reliability already published can be interpreted only when one can estimate the sigmas of the groups involved. Correlations for validity of tests also can be roughly corrected for estimated differences in range.[3]

2. The accumulation of such data will render it possible to select a typical range as a basis of reference for the correction of obtained reliability coefficients. For example, it has been suggested that the sigma of Stanford-Binet mental age of unselected 12 year old children may be taken as the standard range to which reported coefficients may be corrected.[4] A similar standard sigma of intelligence quotients could be selected. Obtained coefficients might well be so corrected by persons reporting them.

---

[1] The writer is especially indebted for suggestions and advice, to Dr. John P. Herring, Director of the Bureau of Research.

[2] Formula $\frac{\sigma}{\Sigma} = \frac{\sqrt{1-R}}{\sqrt{1-r}}$, from Kelley, T. L.: "Statistical Method," p. 222.

[3] Ibid., p. 228ff.

[4] Herring, J. P.: Avery's Comparison of the Stanford and Herring Revisions; *Journal Educational Psychology*, Sept., 1924 (in a footnote).

TABLE I.—(A) SIGMA OF MENTAL AGES IN GRADE GROUPS

(Throughout Table I number in parenthesis indicates number of pupils tested)

Data from group number	Grade: Kdgn.	I	II	III	IV	V	VI	VII	VIIA or VIIB	VIII	IX
1. Herring-Binet	....	4.56 (29)	7.42 (21)	8.22 (27)	14.42 (33)	15.18 (30)					
2. Stanford-Binet	....	....	9.43 (25)								
			8.09 (29								
			7.85 (24)								
			7.81 (28)								
			6.06 (21)								
			10.39 (26)								
3. Haggerty Delta I	....	....	9.52 (16)	9.97 (19)							
			12.61 (19)	7.16 (12)							
4. NIT (A)	....	....	.....	....	.....	21.84 (16)	17.23 (18)	24.74 (19)			
5. NIT (A)	....	....	.....	....	.....	14.59 (14)	16.03 (19)	13.78 (15)			
6. NIT (A)	....	....	.....	....	.....	14.76 (25)	15.48 (16)	15.66 (10)			
7. NIT (A and B)	....	....	.....	....	7.90 (13)	7.23 (17)	6.32 (12)	11.77 (12)			

8. NIT (A and B)	....	....	.....	....	5.21 (19) 6.75 (22)	15.05 (27) 8.27 (19)	13.06 (24) 9.75 (28)					
9. NIT (A)	....	....	.....	....	10.98 (30)	.....	12.61 (14)					
10. Herring-Binet	....	7.95 (27)	4.50 (28)	6.15 (23)	6.80 (23)	8.32 (24)	11.25 (25)	.....	5.39 (22) 8.25 (20)	7.87 (25)	17.89 (33)	
11. Herring-Binet	6.18 (41)	9.44 (30)	5.61 (26)	8.00 (24)	9.66 (22)	8.84 (25)	12.77 (26)	.....	13.47 (28) 8.43 (19)	8.49 (26)	18.79 (34)	
12. Stanford-Binet	....	....	.....	....	8.9 (32)	10.5 (27)	16.7 (27)					
13. New Jersey Composite	....	....	.....	....	.....	19.15 (c10,000)						
14. Thorndike Non-language	....	....	.....	....	.....	.....	10.5 (361 boys)					
15. Pressey Cross-out	....	....	.....	....	20.07 (971)	20.51 (1042)	20.07 (1057)	9.34 (998)	.....	20.51 (723)	36.63 (303)	Grade X 19.19 (184) Grade XI 18.32 (151) Grade XII 17.00 (74)
16. Haggerty Delta II	....	....	.....	....	20.19 (950)	23.47 (1020)	21.78 (916)	21.63 (737)	.....	20.85 (689)	18.42 (173)	

TABLE I.—(B) SIGMA OF MENTAL AGES IN AGE GROUPS.—*Continued*

Data from group number	Age in Years 6	7	8	9	10	11	12	13	14	15	16	Adult
17. Stanford-Binet	.....	.....	.....	.....	.....	.....	25.11 (116)					
18. Stanford-Binet	14.83 (?)						29.65 (c90)					
19. Stanford-Binet	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	31.8 (653)
20. Stanford-Binet	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	45.52 (c2400)
21. Army Alpha	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	40.32 (c2400)
22. NIT (A and B)	.....	.....	.....	14.65 (c175)	18.15 (c175)	19.35 (c175)	22.65 (c175)	23.75 (c175)	27.50 (c175)			
23. NIT (A)	.....	.....	.....	.....	.....	.....	26.56 (3856)					
24. Haggerty Delta II	.....	.....	.....	.....	.....	.....	30.92 (787)					
25. New Jersey Composite	.....	.....	.....	.....	.....	23.37 (91)	18.27 (94)					
26. Pressey Cross-out	.....	.....	.....	.....	.....	32.82 (1034)	29.01 (1038)	31.50 (960)	19.17 (297)	26.95 (226)	23.73 (162)	

TABLE I.—(C) SIGMA OF EDUCATIONAL AGES IN GRADE GROUPS.—*Continued*

Data from group number	Grade II	III	IV	V	VI	VII	VIII	IX
27. Stanford Achievement	....	9.86 (?)	8.96 (?)	11.46 (?)	11.91 (?)	15.07 (?)	14.52 (?)	15.71 (?)
28. Stanford Achievement	....	.....	.....	.....	.....	.....	13.22 (96)	
29. Stanford Achievement	8.62 (33)	11.76 (31)	10.08 (31)	.....	8.33 (15)			
30. A composite	8.55 (24)	5.91 (21)	9.46 (24)	8.39 (24)	9.77 (26)	10.45 (22) 8.18 (20)	11.85 (24)	15.14 (33)

TABLE I.—(D) SIGMA OF EDUCATIONAL AGES IN AGE GROUPS.—*Continued*

	Age in Years 7	8	9	10	11	12	13	14	15
31. Stanford Achievement	12.71 (?)	18.98	18.01	21.77	19.59	21.47	24.31	25.29	25.29

TABLE I.—(E) SIGMA OF INTELLIGENCE QUOTIENTS IN GRADE GROUPS.—*Continued*

Data from group number.....	Grade I	II	III	IV	V	VI	VII	VIII	IX
10. Herring-Binet.....	7.47 (28)	13.32 (28)	13.12 (23)	13.30 (24)	14.06 (24)	17.30 (26)	9.70 (42)	13.20 (25)	16.86 (33)
2. Stanford-Binet.....	....	13.60 (25)							
		10.40 (29)							
		14.29 (24)							
	—	10.60 (28)							
		10.01 (21)							
		12.26 (26)							
32. Stanford-Binet.....	....	.....	.....	.....	.....	.....	.....	.....	15.6 (107)
33. Stanford-Binet.....	....	.....	.....	.....	.....	17. (17)	15. (226)	17. (39)	
34. Terman Group Test.....	....	.....	.....	.....	.....	21. (17)	14. (171)	16.	
							14. (226)	(169)	
							14. (171)	19. (39)	
							18.1 (37)		
							18.5 (32)		

TABLE I.—(F) SIGMA OF INTELLIGENCE QUOTIENTS IN AGE GROUPS.—*Continued*

	12 years	
17. Stanford-Binet.....	16.9 (116)	
35.	17.9	(By computation from mean sigma of MA of 12-year-old age groups)

3. A knowledge of the sigma of mental or of educational ages to be expected in various groups would make it possible to predict the reliability coefficient of a test at a certain grade or age level (assuming that the test is an equally reliable measure at different levels and that the reliability in a given range is known).

4. The accumulation of data as to sigmas of test ages for various subjects and abilities as measured by various tests will ultimately furnish a basis for judging the relative dispersion (within a typical group) of general abilities such as mental age, of slightly more circumscribed abilities such as educational age, of somewhat narrower abilities such as reading or arithmetic age, and of abilities usually considered still more specialized such as spelling age or musical or mechanical aptitude in terms of test age. Thus it may be that the sigma of arithmetic ages of an unselected 12 year old group is less than is the sigma of such a group in mental age, even when the two functions are measured with equal reliability. In other words, different functions probably correlate differently with chronological age.

5. An indication of the average homogeneity and of the attainable homogeneity of grade groups with respect to the various abilities (as measured by tests differing both in reliability and in correlation with age) would set up a desirable objective for reclassification. The gradual reduction of mean sigmas of grade groups in a school system to a point well below the comparable means of other similar systems may turn out to be one index of the efficiency of the supervision.[1]

6. Knowledge of the relative values to be expected for sigmas of various school subjects within a grade as measured by certain tests would make it possible to express the sigmas of a class group in each subject in units of the mean sigma of such a group in that subject. Thus the subjects most in need of attention from this point of view could be justly determined.

Some Sigmas of Typical Groups

Until the significance of varying locations, types of school, subjects tests, etc., is known, it is necessary to report data as to the source of

[1] Such comparisons, if between grades in which different tests have been used, should involve a correction for any differences which may exist in the correlations of the tests with chronological age. It should be determined whether or not, for tests A and B, $\frac{\sigma_A}{\sigma_B}$ in a grade group is equal to $\frac{\sigma_A}{\sigma_B}$ in an age group at the corresponding level, or whether some other means of correction should be used.

(By computation from mean sigma of MA of 12-year-old age groups)
(116)
17.9
35.

each sigma collected. It should be kept in mind that sigmas vary with the reliability of the test as well as with the true dispersion of the group in the function measured; and that, in age groups, the dispersion itself bears a functional relationship to the correlation of the test with age and to the degree of imperfection of the sampling.

In Table I, the group numbers in the left hand column refer to the source of the data as given in the note at the end of this article.

TABLE II.—MEAN SIGMAS. (A) SIGMA OF MENTAL AGE IN GRADES

(N indicates number of sigmas entering into each average)

1. Little previous reclassification (sources 1 to 9 inclusive as numbered in Table I).

Grade	Kdgn	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Mean sigma	....	4.56	8.79	8.45	8.86	13.84	12.92	15.79					
N	....	(1)	(9)	(3)	(5)	(7)	(7)	(5)					

2. Much previous reclassification (sources 10 to 12 inclusive as numbered in Table I).

Mean sigma	6.18	8.69	5.05	7.07	8.45	9.33	13.57	9.09	8.49	18.79			
N	(1)	(2)	(2)	(2)	(3)	(3)	(3)	(3)	(1)	(1)			

3. Groups (1) and (2) combined (sources 1 to 12 inclusive as numbered in Table I).

Mean sigma	6.18	7.32	8.12	7.90	8.75	12.45	13.12	13.28	8.49	18.71			
N	(1)	(3)	(11)	(5)	(8)	(10)	(10)	(8)	(1)	(1)			

4. Composite groups (several class groups combined at each grade. Sources 13 to 16 inclusive as numbered in Table I).

Mean sigma	....	....	....	....	20.08	21.04	17.45	20.48	20.68	27.52	19.19	18.32	17.00
N	....	....	....	....	(2)	(3)	(3)	(2)	(2)	(2)	(1)	(1)	(1)

TABLE II.—(B) SIGMA OF EDUCATIONAL AGE IN GRADES

1. Single grade groups.

Grade	II	III	IV	V	VI	VII	VIII	IX
Mean sigma	8.59	8.84	9.77	8.39	9.05	9.32	11.85	15.14
N	(2)	(2)	(2)	(1)	(2)	(2)	(1)	(1)

2. Composite groups. (Several class groups combined at each grade.)

Mean sigma	....	9.86	8.91	11.46	11.91	15.07	13.92	15.71
N	....	(1)	(1)	(1)	(1)	(1)	(2)	(1)

TABLE II.—(C) SIGMA OF MENTAL AGE IN AGE GROUPS

Age	6	9	10	11	12	13	14	15	16	Army white draft	
										General	English speaking
Mean sigma	14.83	14.65	18.15	25.18	26.02	27.62	28.33	26.95	23.73	42.92	31.80
N	(1)	(1)	(1)	(3)	(7)	(2)	(2)	(1)	(1)	(2)	(1)

TABLE II.—(D) SIGMA OF EDUCATIONAL AGE IN AGE GROUPS

Age	7	8	9	10	11	12	13	14	15
Sigma	12.71	18.98	18.01	21.77	19.59	21.47	24.31	25.29	25.29
N	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)

TABLE II.—(E) SIGMA OF IQ IN GRADE GROUPS

Grade	I	II	III	IV	V	VI	VII	VIII	IX
Mean sigma	7.43	11.92	13.12	13.30	14.06	18.43	14.76	16.30	16.23
N	(2)	(7)	(1)	(1)	(1)	(3)	(7)	(4)	(2)

CONCLUSIONS

Mean sigmas based on different tests at different levels are only approximately comparable due to the differences in correlation of the tests with age. Yet the sigmas of tests in general are more comparable with these mean sigmas than they are with each other. It is desirable, of course, to continue the collection of sigmas as here begun in order to increase the reliability of the conclusions. The following trends are suggested:

1. The sigmas of mental age in grade groups increase from the lower to the higher grades in schools where little previous reclassification according to standardized test results has taken place, but no such trend is apparent in more carefully reclassified schools.[1] The relatively low sigmas of mental age (and probably of educational age) in the primary grades of the typical (unreclassified) school may be

[1] The high sigma of MA (18.71 months) reported in Grade IX in a reclassified school is due in part at least to the fact that there is no outlet above the ninth grade in this school.

a partial explanation of the apparently low reliability and validity correlations usually found in primary grades.

2. The sigmas of educational age in grade groups tend to be lower than the corresponding sigmas of mental age, especially in middle and upper grades. This trend seems logical as grading is ordinarily done on a basis of educational rather than intelligence level.

3. Composite grade groups (*e.g.*, several fifth grades combined) show much higher sigmas of mental age than do single grades. Such sigmas often approach those of the younger age groups; for example, those of 9- or 10-year-old children.

4. Concerning the sigmas of mental ages in a series of grades taken as a single group, the following facts may be stated:

(*a*) In unreclassified schools the sigmas of mental age of two successive grades taken as a single group average 119 per cent of the average sigmas of the two grades considered separately. (Eleven such sigmas of pairs of grades are averaged in this case.)

(*b*) In reclassified schools the corresponding figure is 129 per cent (average of seven sigmas).

(*c*) Sigmas of mental ages of three successive grades are 149 per cent (average of two sigmas) and 155 per cent (average of three sigmas), for unreclassified and reclassified schools, respectively.

(*d*) The sigmas of mental ages of four grades taken as a single group approximate those of a 12-year-old age group, being in three cases 28.63 months, 23.11 months, and 26.36 months, respectively.

5. The sigma of mental ages in age groups appears to increase fairly regularly from the 6-year group to the 15-year group. This increase, in the present data, amounts to nearly 100 per cent. Of course, most age groups reported upon are to some extent selected.

6. The sigmas of educational age of age groups show a similar tendency to increase from age to age but are uniformly somewhat lower than sigmas of mental age, although the number reported is small. (The tendency for the accomplishment ratios of dull children to be above 1.00 and for the bright children to be below 1.00 would lead one to expect the sigma of educational age in an age group to be smaller than the sigma of mental age in the same group.)

7. The mean sigma of mental ages of seven 12-year-old age groups is 26.02 months according to these data. This age group is of sufficient

importance to merit a somewhat closer analysis. The sigmas included in this mean may be read from Table IB and the source of each sigma may be ascertained by its group number from the note at the end of this article. Let us examine these sigmas:

A sigma of 25.11 months is reported from the Stanford Revision of the Binet Test (Group 17). Dr. Herring reports a sigma of 25.22 months from the Herring-Binet in the same group and a reliability for these tests in this group of .987 (Herring and Stanford mental ages correlated). Averaging these sigmas and correcting for the unreliability of the tests according to the formula $\sigma_{tru} = \sigma_{obt}\sqrt{r_{11}}$ we have $\sigma_{tru} = 25.01$ months for this group.

A sigma of 18.27 is reported from the New Jersey composite test (Group 25). This group does not include 12-year-olds above Grade VIII and is therefore somewhat restricted at the high levels. We should probably not attempt to make correction for this restriction but should throw out this sigma for the present.

A sigma of 22.65 is given for the National Intelligence Test. An examination of the source of this sigma (Group 22) will show that the sigma is an approximation only and should therefore be discarded from this more critical list.

A sigma of 29.65 is reported from the Stanford-Binet (Group 18). If we assume that the reliability coefficient of this test in this group is .96 (which is certainly attainable by careful testing), we may correct this obtained sigma to a true sigma of 29.06 months.

A sigma of 29.01 is reported from the Pressey Cross-out Test (Group 26). In the reference cited for this group a reliability coefficient for 86 Grade VII children is reported as $r_{1I} = .82$. Taking 13.25 as the mean sigma of Grade VII (Table II(A)) and increasing this reliability coefficient to what it would be in a group having a mean sigma of 27 months[1] we obtain a corrected coefficient of .96. If we correct the obtained sigma of 29.01 months for a reliability of test of .96 we find a true sigma of 28.42 months.

A sigma of 26.56 is reported from the National Intelligence Test, Scale A (Group 23). A reliability coefficient of .922 (form with form) has been reported for this test in a group having a sigma of mental age of 23.9 months.[2] Correcting this coefficient to what it would be in a group having a mean sigma of mental age of 27 months, we obtain a

[1] Kelley, p. 222 (27 months is taken to be the approximate mean sigma of 12-year-olds for the purpose of this correction).

[2] Symonds, P. M.: *Journal Educational Psychology*, Oct., 1924, p. 424.

corrected coefficient of .939 as in a 12-year-old age group. Now correcting the sigma of 26.56 on the basis of this reliability we obtain a true sigma of 25.73 months.

A sigma of 30.92 is reported from Haggerty Intelligence Examination Delta II (Group 24). If we assume that this test has a reliability of .90 in a 12-year-old age group (which seems conservative), we may obtain a corrected sigma of 29.99 months.

The mean of these five corrected sigmas, weighted according to the number of scores entering into each, is 26.81 months which is, from these data, the closest possible approximation to the true sigma of mental age of a 12-year-old group in the functions measured by these tests.

8. The sigma of mental age of an unselected adult group appears to be about 60 per cent greater than that of a 12-year-old group.

9. The sigmas of intelligence quotients in grade groups show little tendency to increase from grade to grade and are as high in the middle and upper grade groups as for a 12-year-old age group. It is probable that a sigma of IQ of about 17.5 points may be expected on the average in a group not selected on the basis of IQ. Correlations involving intelligence quotients are not directly comparable with those involving mental or educational ages, even in the same group.

NOTE.—The following are statements of the sources of the data in this article. Numbers correspond to those in Table I.

1. Herring Revision of the Binet Test given in the Training School of a Pennsylvania State Normal School. Little reclassification on the basis of standardized test results had preceded this testing. From Herring and Wilner: "Manual for Measuring a School;" World Book Co., p. 27.

2. Stanford-Binet Test. City School, New York. Probably little previous reclassification. From Graves, Katherine B.: Specialized Training in Tests for General Intelligence; *Teachers College Contributions to Education*, No. 143 (1924). Children to the number indicated selected alphabetically from each room.

3. Haggerty Delta I Intelligence Examination. Normal Training School, New Jersey. Little previous reclassification. Author's data.

4, 5, and 6. National Intelligence Test, Scale A, given in three five-room rural schools, New Jersey. No previous reclassification. Author's data.

7. National Intelligence Test, Scales A and B combined, given in a six-room rural school, New Jersey. No previous reclassification. Author's data.

8. National Intelligence Test, Scales A and B combined, given in the Training School of a New Jersey State Normal School. Little previous reclassification by standardized tests. Author's data.

9. National Intelligence Test Scale A. A County Orphans' Home, Connecticut. Little previous reclassification. Author's data.

10. Same as group 1 above except that these sigmas apply to the grade groups immediately after reclassification on the basis of intelligence and achievement tests.

11. Same as group 10 except nine months after reclassification.

12. Stanford-Binet Test. In a city school where much previous reclassification had been done, New York. From Herring, J. P.: "The Herring Revision of the Binet Simon Tests and Verbal and Abstract Elements in Intelligence Examinations:" World Book Co., 1924, pp. 48–50.

13. New Jersey Composite Test—Sub-tests 1 to 3, and 8 to 10; (comprising those tests considered to be intelligence tests). Public schools of New Jersey. Derived by the writer from unpublished data on sigmas of scores in the hands of the committee on this test.

14. Thorndike Non-language Intelligence Examination. Boys from average and superior social groups. Derived by the writer from data given in Thorndike, E. L.: A Standardized Group Examination of Intelligence Independent of Language; *Journal of Applied Psychology*, Mar., 1919, p. 13.

15. Pressey Cross-out Test. Composite groups. Derived by the writer from data given in Pressey, S. L.: A Brief Group Scale of Intelligence for Use in School Surveys; *Journal of Educational Psychology*, Feb., 1920, p. 89.

16. Haggerty Intelligence Examination. Delta II. Composite groups. Idaho. Derived by the writer from data given in Madsen, I.N.: Intelligence as a Factor in School Progress; *School and Society*, March 11, 1922, p. 283.

17. Stanford-Binet Test. A group of 12-year-old public school children possessing to a notable degree the characteristics of an "unselected" group. Pennsylvania. Herring, J. P.: Reliability of the Stanford and Herring Revisions of the Binet-Simon Tests; *Journal of Educational Psychology*, April 1924, 217.

18. Stanford-Binet. Reference same as for 22 below.

19. Stanford-Binet, four tests per year, army men, white, "Born in English-speaking Countries." From *Memoirs of the National Academy of Sciences;* Vol. XV; Government Printing Office, p. 388.

20. Stanford-Binet (four tests per year) white draft. Reference same as 19 above, but p. 634.

21. Army Alpha. Reference same as 20 above.

22. National Intelligence Test, Scales A and B. "Unselected" children. Approximate sigmas derived by the writer from Terman, L. M.: Mental Growth and the IQ; *Journal of Educational Psychology*, Oct., 1921, p. 401. Dr. Terman states that "the sigma of total score of scales A and B remained almost constant at 50 points throughout this entire range" (9 to 12 years).

23. National Intelligence Test, Scale A. Derived by the writer from data given in Supplement 3 to the Manual of Directions accompanying the test. World Book Co.

24. Haggerty Intelligence Examination Delta II. Reference same as for 16 above.

25. New Jersey Composite Test, sub-test 1 to 3, and 8 to 10 (considered to be intelligence tests). All 11- and 12-year-old children in Grades III to VIII inclusive in two city schools. New Jersey. Writer's data.

26. Pressey Cross-out Test. Reference same as for 15 above. An effort was made to secure all school children of the ages specified.

27. Stanford Achievement Test. Composite groups derived by the author from percentile tables given in the manual accompanying the test. California. World Book Co.

28. Stanford Achievement Test. Composite of several eighth grades in a city system where a good deal of previous reclassification has probably been done. Derived by the writer from data given in Kelley, T. L.: A New Method of Determining the Significance of Differences in Intelligence and Achievement Test Scores, *Journal of Educational Psychology*, Sept., 1923, p. 321.

29. Stanford Achievement Test, given in a County Orphans' Home. Little previous reclassification. Connecticut. Writer's data.

30. A composite educational age derived from subject ages gained in Thorndike-McCall Reading Test, Woody Arithmetic Test Series A, and Nassau County Supplement to the Hillegas English Composition Scale. Testing in the Training School of a State Normal School shortly after careful reclassification of the pupils. Pennsylvania. Unpublished data furnished by Dr. John P. Herring.

31. Stanford Achievement Test. Derived by the writer from the PE of scores and reliability coefficients (given in the manual accompanying the test) according to the formula: $\sigma_{\infty 1}=\sigma_1 \sqrt{r_{11} - r_{11}^2}$.

32. Stanford-Binet Test. Derived by the author from data given in Proctor, Wm. M.: The Use of Intelligence Tests in the Educational Guidance of High School Pupils, *School and Society:* Oct. 26, 1918, p. 502.

33. Stanford-Binet Test. Composite groups. Derived by the writer from data in Association Notes on Educational Research; *Journal of Educational Research*, May, 1923, p. 458.

34. Terman Group Test of Mental Ability. Reference same as for 33 above.

35. Assuming the sigma of mental age of 12-year-olds to be 26.81 months and their mean MA to be 150 months, the sigma of IQ of a 12-year-old group would be 17.9 points.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY C. O. MATHEWS

Intelligence Testing

The Improvement of Mental Measurements. Edward L. Thorndike. Journal of Educational Research, Jan., 1925, pp. 1–11. Discussion of the advisability of separate units to measure level, range, speed and method of intellectual tasks.

The Nature of Intellect. Edward L. Thorndike. The Educational Record, Jan., 1925, pp. 3–12. Data supporting the hypothesis that "the higher forms of intellectual operation are identical with mere association or connection forming, depending upon the same sort of physiological connections but requiring *many more of them.*"

The Ohio Literary Test. Violet H. Foster and Henry H. Goddard. Pedagogical Seminary and Journal of Genetic Psychology, Dec., 1924, pp. 340–351. History, standardization and norms for this test. Correlates highly with intelligence tests. Useful for individuals or groups.

The Intelligence of White Foreign Born Criminals. Carl Murchison. Pedagogical Seminary and Journal of Genetic Psychology, Dec., 1924, pp. 297–307. A study of racial differences in intelligence among foreign born criminals in the state prisons of Ohio, Illinois, New Jersey and Maryland.

The Relationship between Three Tests of Imagination; and Their Correlation with Intelligence. John A. McGeoch. The Journal of Applied Psychology, Dec., 1924, pp. 439–443. A study of relationships between results on the ink-blot, word-building and linguistic invention tests, and their correlations with Army Alpha results.

A Survey of the Intelligence and Environment of School Children. Tonan Fukuda. The American Journal of Psychology, Jan., 1925, pp. 124–139. The relation of intelligence to nationality, school training, and environment shown in a survey of the Washington School, Evanston, Illinois.

Achievement Testing

Motivating the Course in Tests and Measurements for the Teacher-in-training. Emma B. Grant. The Journal of Educational Method, Jan., 1926, pp. 190–194. Procedure used in training teachers in a large city training school.

A Comparative Study of Certain Reading Tests—II. Sister M. Kathleen. The Catholic Educational Review, Jan., 1925, pp. 33–42. A comparison of Courtis, Thorndike-McCall, Stone Series, Chapman-Cook and Pressey Vocabulary Tests.

The Response of Epileptic Children to Mental and Educational Tests. J. Tylor Fox. The British Journal of Medical Psychology, Nov., 1924, pp. 235–248.

The results of 14 tests, intelligence, reading, spelling and arithmetic, with 150 epileptic school children in England.

An Effort to Measure Typing Efficiency. William A. Cook. Journal of Educational Research, Jan., 1925, pp. 49–59. A practical method of measuring typing which can be used by any teacher with materials he may select.

Correcting Examination Papers. Retha E. Breeze. The School Review, Jan., 1925, pp. 57–61. Gives evidence to show that pupils are benefited to a large extent by having errors on examinations pointed out to them.

Achievement in High School Chemistry—an Examination of Subject-matter. S. R. Powers. School Science and Mathematics, Jan., 1925, pp. 53–61. Results of measures of achievement in chemistry at the close of high school courses and at the beginning of college courses.

An Analysis of Several Well Known Tests. H. E. Garrett and V. W. Lemmon. The Journal of Applied Psychology, Dec., 1924, pp. 424–438. Analyses of the A and the a-t cancellation tests and the Color Naming Test by the method of partial correlation.

Learning and Guidance

Directed Drill in the Comprehension of Verbal Problems in Arithmetic. Harry A. Greene. Journal of Educational Research, Jan., 1925, pp. 33–40. An experiment to determine the value of drill in training to recognize the process involved in solving verbal problems.

A Summary of Some Significant Conclusions Reached by Investigators Relative to Arithmetic. J. C. Brown. The Elementary School Journal, Jan., 1925, pp. 346–357. A summary of investigations since Rice's study in 1902, with excellent bibliographies on scientific studies and on tests and diagnosis in arithmetic.

An Experimental Study in Patterns of Thought. Garry C. Myers. Pedagogical Seminary and Journal of Genetic Psychology, Dec., 1924, pp. 352–358. Shows that solving simple problems in algebra in elementary grades is dependent upon becoming familiar with patterns of thought.

A Physiological-behavioristic Description of Thinking. J. F. Dashiell. Psychological Review, Jan., 1925, pp. 54–73. A description of thinking "as a human function that involves some substituting of implicit for explicit reactions."

A Study of Individual Differences in Learning. Margaret Kincaid. Psychological Review, Jan., 1925, pp. 34–53. The application of seven measures of differences to 24 learning experiments concluding "that differences generally decrease with practice."

How Two Hundred and Fifty-eight Junior College Women Study. Jessie Allen Charters. Journal of Educational Research, Jan., 1925, pp. 41–48. Four methods of study were investigated, namely: Studying a textbook assignment, memorizing, distraction and concentration, and note-making.

A Laboratory Study of Fear: The Case of Peter. Mary Cover Jones. Pedagogical Seminary and Journal of Genetic Psychology. Dec., 1924, pp. 308–315. A case study showing that a fear response can be removed by laboratory methods and that this "unconditioning" may spread to other stimuli.

An Experimental Investigation of the Influence of Certain Weather Conditions upon Short Periods of Mental Work. J. P. Bethel. The American Journal of Psychology, Jan., 1925, pp. 102–112. The effects of pressure, humidity and

temperature upon visual memory of syllables and auditory memory of numbers. Also includes a historical survey of related investigations.

Original Nature and the Pre-school Years. G. E. Johnson. Progressive Education, Jan., Feb., Mar., 1925, pp. 5–7. Some practical suggestions for the guidance of pre-school children.

Maladjustment among College Students. Zoe Emily Leatherman and Edgar A. Doll. The Journal of Applied Psychology, Dec., 1924, pp. 390–410. An analysis of causes of maladjustment and 14 case studies from the Psychological Clinic of Ohio State University.

The Need of a Counseling Program in Secondary Schools. Virgil E. Dickson. Journal of Educational Research, Jan., 1925, pp. 12–16. Data showing variation in chronological and merital ages and IQ's of beginning high school pupils and their need of guidance.

Miscellaneous

Language as a Higher Form of Reaction. Charles H. Judd. The Elementary School Journal, Jan., 1925, pp. 335–345. The fourth of a series of articles on educational psychology. The importance of language is pointed out as a means of connecting the individual with past accumulations of knowledge.

Spurious Correlations in Psychology. James Quinter Holsopple. Pedagogical Seminary and Journal of Genetic Psychology, Dec., 1924, pp. 399–404. The correlation of variables containing common elements produces spurious correlations. Discusses desirability of such correlations.

Classification of Personalities. Edmund S. Conklin. Pedagogical Seminary and Journal of Genetic Psychology, Dec., 1925, pp. 316–332. A scheme for classifying personalities according to their positions on the extra-introversion, intelligence, noetic and moral scales, recognizing three diatheses: The balanced, the cyclothymic and the epileptic.

The Distribution of Marks in High Schools of Varying Size. D. A. Emerson. School and Society, Jan. 10, 1925, pp. 54–56. In a study of marks in 30 high schools the smaller schools gave a larger proportion of high marks.

The Vocabularies of High School Science Textbooks. S. R. Powers. Teachers College Record, Jan., 1925, pp. 368–382. Word counts of texts in General Science, Biology and Chemistry.

The Content of Ordinary Reading. H. L. Donovan. The Elementary School Journal, Jan., 1925, pp. 370–375. A study of the space alloted to subjects in the *Chicago Daily Tribune* and *Literary Digest.*

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY JOHN HOCKETT

The Lincoln School of Teachers College

A Plea for Better Examinations

Improvement of the Written Examination, by G. M. Ruch. New York: Scott, Foresman & Co., 1924.

The thanks of the whole teaching profession are due to Dr. Ruch for publishing this neat and pointed exposition of the principles and the best practice of examination-making. It should be an early and indispensable addition to the book shelf of every teacher who wishes to learn with as much economy as possible how to improve examinations or who wishes to be abreast of the rapidly developing science of educational measurement. This book will do much to lessen the difference between the rate of development and the rate of practical application of the newer examining methods.

Many teachers have been keenly conscious of the inadequacies of the traditional methods of measurement which have come down to us almost without change from the Middle Ages, but these teachers have not found it as convenient to understand the constructive suggestions that have been offered as to feel the full force of the destructive criticisms that have been poured out in increasing volume against the old methods. It is for this reason that so clear and concise a book as Dr. Ruch's is so timely just now. The well-chosen sample questions, the detailed suggestions for actual examination construction, the "Questions and Problems" at the end of each chapter, and the absence of involved and highly technical details that have made many books on measurement of the last few years very difficult, if not impossible for the average teacher, all contribute to make this book as valuable to the teacher of beginning classes in measurement as to the teacher in the field.

In Chap. I the author analyzes the functions of the written examination. The issue between the claims of pedagogy and of meas-

urement are clearly drawn. That the old type examinations have little measurement value is as clear as that the objective tests do have considerable value as measuring instruments; that neither old nor new type tests solve the pedagogical problems of all subject matters is equally clear. "The solution would seem to be that a new methodology must be created for language training and measurement apart from the more general examination for factual mastery." In the meantime, we cannot afford to reject good measuring instruments because they are suspected of pedagogical weaknesses; certainly the objective examinations cannot be rejected in favor of a type of examination which has neither measuremental nor pedagogical values which are above question.

Chapters II, III and IV give in less than 100 pages lucid expositions of the Criteria of a Good Examination, Sources of Error in Written Examinations, and Types and Construction of the Newer Objective Examinations. Chapter V presents the results of some important experimental studies of several types of objective examinations, while Chap. VI is concerned with several statistical considerations related to examination technique in general. In addition to many illustrations of objective questions in Chap. IV, an appendix reproduces in full the Ruch High School Content Examination, an objective examination in first-year French, one in first-year Latin and one in Manual Training.

BEN D. WOOD.

Columbia University.

FURTHER ANALYSIS OF INTELLIGENCE TESTS

Content and Form in Tests of Intelligence, by E. M. Bailor. New York: Teachers College, Columbia University Contributions to Education, No. 162, 1924. Pp. 74.

The purpose of this study was to determine the influence of differences in content and form on test results in tests of intelligence. Content here refers to the kinds of data (verbal, numerical or spatial) and form to the "objective mold in which the test is cast" (completion tests, analogies tests, generalizations tests, and a composite of five tests of different objective composition).

Thorndike's tests from his "Mental Discipline in High School Studies" were used for the major part of the study. These were 15

tests designated as tests of selective and relational thinking and generalization and organization.

The subjects were students of the ninth, tenth and eleventh grades of a city of the Middle West whose population is mostly native whites. Slightly more than 1000 boys' records were used. Age control groups and time control groups were formed.

The author finds that differences in the relative standing of pupils occur when they are given tests having differences either in form or content. He finds no adequate evidence to indicate that the results of tests differing in content and in form are duplicate and independent measures of a common mental function, such as Spearman claimed. Correlations throughout were positive.

A statistical contribution of this study is the development of the predictive index (PI), which is the arithmetical complement of the coefficient of alienation described by Kelley. It is suggested as a measure of the significance of a correlation coefficient from the viewpoint of prediction. Its preference over k lies in the fact that it is a positive statement of the ratio. BETH WELLMAN.

AN OBJECTIVE CONTRIBUTION TO THE GRAMMAR QUESTION

How Much English Grammar? by Martin J. Stormzand and M. V. O'Shea. Warwick & York, Inc., Baltimore, 1924. Pp. 224.

This is the most significant report of educational research on the grammar problem since the Charters-Miller bulletin of 1915. Dr. Charters and Miss Miller in their study postulated error frequency as a criterion for the selection of grammar content; Dr. Stormzand and Dr. O'Shea have sought to add to that the criterion of use frequency. We need still a third, namely, that of experimentally determined results; but the two above are basic.

The relative importance of formal grammatical concepts from the point of view of use frequency has been determined in this study by the detailed analysis of "ten thousand sentences selected from material that represented all grades of current usage." A great number of summary tables give the percentages within categories, or the sentence ratios, of the kinds of sentences, the kinds of clauses, phrases, and non-modal verb forms, the eight parts of speech and their respective classification types and inflectional forms. These percentages and

ratios are further differentiated for adults; university freshmen and upper classmen; high school seniors, juniors, sophomores, and freshmen; and pupils in each of the grades from the fourth to the eighth. Other differentiations of the data recognize such groupings as business letters, certain journalists, women's letters, specific authors, local newspaper news notes, the Associated Press, newspaper editorials, conversation in novels.

An interesting, if not wholly satisfying, effort is made to compare and combine the results of this analysis with the data from the few authentic error studies in print and with the authors' own error count of the school compositions in their material. The significant outcome of this is a table of *error quotients*, giving for 34 items "the number of errors that may be expected in school compositions in 1000 chances." The book closes with a comparison of "the relative importance given various topics in the grammar textbooks in general use (10 such named) with the values we have established as a result of our studies in use and error." Recommendations are made for the revision of grammar texts, which, if followed, will hasten the excision of traditional matter and will cause a considerable shift in emphasis among the topics that may properly be retained.

Defects the study has, but they seem to be due inevitably to the overwhelming amount of work that such analysis involves, and to the meagerness of reported research that might support, supplement, and check the investigators in their procedure. Further studies of a similar kind are desirable—studies undertaken with much greater resources and carried through with still greater refinement of technique—but the general conclusions of the present one are not likely to be superseded.

M. H. Willing.

Guidance in Presenting Facts

Graphic Methods in Education, by J. Harold Williams. Boston: Houghton Mifflin Co., 1924. Pp. XVII + 319.

As the theory and practice of education come to rest on a more scientific basis, as measurement and numerical representations assume increasingly greater importance in our educational thinking, there develops a need for wider knowledge of the best means of presenting numerical data in forms that will attract the attention of the reader

and will result in the most rapid, yet thorough comprehension of the facts presented. The purpose of the present book is to convince the school man of the value and possibilities of graphic presentation and guide him in the construction or evaluation of the most effective types of charts for various purposes.

One chapter, devoted to the mechanics of the preparation of charts, is designed to acquaint the novice with the instruments and materials needed for chart making and to instruct him in simple drawing and lettering. Fifteen chapters are devoted to the description and critical evaluation of as many types of charts. Mention of some of these types will make evident the scheme of classification. They include Bar Representations, Circle Representations, Curves, Frequency Surfaces, Maps, Picture Graphs, Architectural Diagrams, and many more. Each chapter is copiously illustrated with charts portraying educational data of many kinds. There are, in fact, 148 illustrative charts in the book.

Graphic Methods in Education may be used as a textbook in graduate courses in education. It is, in reality, an outgrowth of a course in "Graphic Methods of Presenting Facts" given in Stanford University by the author and others. The book will, however, also make its appeal to the administrator in the field, to the psychologist, and to all persons, in fact, who have occasion to present numerical or statistical data in the most intelligible form. The book is well planned and well written and should prove of distinct value to the profession. J. H.

Proposal of a New Technique for Formulating a Teacher-training Policy

Cost of Training Teachers, by Homer E. Cooper. Baltimore: Warwick & York, 1924. Pp. 112.

In view of the pressure for economy in public expenditures and the vicissitudes of teacher supply during the last few years, there is an increasing demand that the states re-determine their policies of teacher-training on a more scientific basis. This investigation is such an attempt in devising a method whereby a comprehensive body of facts and technique are organized as guides by which to determine upon a teacher-training policy. This is in sharp contrast to the all too prevalent method of personal judgment, concerned with the demands

of temporary expediency largely and subject to all kinds of political intrigue as a consequence.

The study was conducted in connection with the Educational Finance Inquiry and deals with the situation in the state of New York; however, it is the contention of the author that the basic elements in the procedure cannot differ greatly in making such a study in other states.

Such problems as standards of preparation of teachers, status of teaching force as to training, adequacy of existing policy and facilities for teacher-training, teacher turnover, replacement of sub-standard teachers, etc., are subjected to detailed statistical analysis on the basis of which the author advances specific recommendations for their reconsideration in policy making. It is interesting to note that he has ventured formulas, such as that for computing the average length of service rendered, in areas where rough approximations have been the rule, hitherto. There is a profusion of interesting tables and graphs and an appendix proposing a revised form for reporting certificates that would enable the certificating system of any state to ascertain economically the amount of service rendered under any certificate.

This investigation should be of prime significance to those responsible for the administrative aspects of teacher-training and who are in a position to examine critically the validity of the technique and interpretations by conducting a similar statistical study in a somewhat different setting. J. G. KUDERNA.